BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 March 2003 (27.03.2003)

PCT

(10) International Publication Number
WO 03/024199 A2

(51) International Patent Classification[7]: A01H

(21) International Application Number: PCT/US02/30156

(22) International Filing Date:
23 September 2002 (23.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/323,961 21 September 2001 (21.09.2001) US
60/324,001 21 September 2001 (21.09.2001) US

(71) Applicant *(for all designated States except US)*: AVIGENICS, INC. [US/US]; 111 Riverbend Road, Athens, GA 30605 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: RAPP, Jeffrey, C. [US/US]; 265 Pinewood Circle, Athens, GA 30606 (US).

SUTRAVE, Pramod [AU/US]; 101 Ashley Circle, Apartment 3, Athens, GA 30605 (US).

(74) Agents: ANTLER, Adrian, M. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: PRODUCTION OF TRANSGENIC AVIANS USING SPERM-MEDIATED TRANSFECTION

SEQ ID NO: 6

```
TGCCGCCTTC TTTGATATTC ACTCTGTTGT ATTTCATCTC TTCTTGCCGA TGAAAGGATA 60
TAACAGTCTG TATAACAGTC TGTGAGGAAA TACTTGGTAT TTCTTCTGAT CAGTGTTTTT 120
ATAAGTAATG TTGAATATTG GATAAGGCTG TGTGTCCTTT GTCTTGGGAG ACAAAGCCCA 180
CAGCAGGTGG TGGTTGGGGT GGTGGCAGCT CAGTGACAGG AGAGGTTTTT TTGCCTGTTT 240
TTTTTTTTTT TTTTTTTTTT AAGTAAGGTG TTCTTTTTTC TTAGTAAATT TTCTACTGGA 300
CTGTATGTTT TGACAGGTCA GAAACATTTC TTCAAAAGAA GAACCTTTTG GAAACTGTAC 360
AGCCCTTTTC TTTCATTCCC TTTTTGCTTT CTGTGCCAAT GCCTTTGGTT CTGATTGCAT 420
TATGGAAAAC GTTGATCGGA ACTTGAGGTT TTTATTTATA GTGTGGCTTG AAAGCTTGGA 480
TAGCTGTTGT TACAGGAGAT ACCTTATTAA GTTTAGGCCA GCTTGATGCT TTATTTTTTC 540
CCTTTGAAGT AGTGAGCGTT CTCTGGTTTT TTTCCTTTGA AACTGGTGAG GCTTAGATTT 600
TTCTAATGGG ATTTTTTACC TGATGATCTA GTTGCATACC CAAATGCTTG TAAATGTTTT 660
CCTAGTTAAC ATGTTGATAA CTTCGGATTT ACATGTTGTA TATACTTGTC ATCTGTGTTT 720
CTAGTAAAAA TATATGGCAT TTATAGAAAT ACGTAATTCC TGATTTCCTT TTTTTTATC 780
TCTATGCTCT GTGTGTACAG GTCAAACAGA CTTCACTCCT ATTTTTATTT ATAGAATTTT 840
ATATGCAGTC TGTGGTTGGT TCTTGTGTTG TAAGGATACA GCCTTAAATT TCCTAGAGCG 900
ATGCTCAGTA AGGCGGGTTG TCACATGGGT TCAAATGTAA AACGGGCACG TTTGGCTGCT 960
GCCTTCCCGA GATCCAGGAC ACTAAACTGC TTCTGCACTG AGGTATAAAT CGCTTCAGAT 1020
CCCAGGGAAG TGCAGATCCA CGTGCATATT CTTAAAGAAG AATGAATACT TTCTAAAATA 1080
TTTTGGCATA GGAAGCAAGC TGCATGGATT TGTTTGGGAC TTAAATTATT TTGGTAACGG 1140
AGTGCATAGG TTTTAAACAC AGTTGCAGCA TGCTAACGAG TCACAGCGTT TATGCAGAAG 1200
TGATGCCTGG ATGCCTGTTG CAGCTGTTTA CGGCACTGCC TTGCAGTGAG CATTGCAGAT 1260
AGGGGTGGGG TGCTTTGTGT CGTGTTCCCA CAGCGTGCCA CACAGCCACC TCCCCGAACA 1320
CATCTCACCT GCTGGGTACT TTTCAAACCA TCTTAGCAGT AGTAGATGAG TTACTATGAA 1380
ACAGAGAAGT TCCTCAGTTG GATATTCTCA TGGGATGTCT TTTTTCCCAT GTTGGGCAAA 1440
GTATGATAAA GCATCTCTAT TTGTAAATTA TGCACTTGTT AGTTCCTGAA TCCTTTCTAT 1500
AGCACCACTT ATTGCAGCAG GTGTAGGCTC TGGTGTGGCC TGTGTCTGTG CTTCAATCTT 1560
TTAAAGCTTC TTTGGAAATA CACTGACTTG ATTGAAGTCT CTTGAAGATA GTAAACAGTA 1620
CTTACCTTTG ATCCCAATGA AATCGAGCAT TTCAGTTGTA AAAGAATTCC GCCTATTCAT 1680
ACCATGTAAT GTAATTTTAC ACCCCCAGTG CTGACACTTT GGAATATATT CAAGTAATAG 1740
ACTTTGGCCT CAGCCTCTTG TGTACTGTAT TTTGTAATAG AAAATATTTT AAACTGTGCA 1800
TATGATTATT ACATTATGAA AGAGACATTC TGCTGATCTT CAAATGTAAG AAAATGAGGA 1860
GTGCGTGTGC TTTTATAAAT ACAAGTGATT GCAAATTAGT GCAGGTGTCC TTAAAAAAAA 1920
AAAAAAAAAG TAATATAAAA AGGACCAGGT GTTTTACAAG TGAAATACAT TCCTATTTGG 1980
TAAACAGTTA CATTTTTATG AAGATTACCA GCGCTGCTGA CTTTCTAAAC ATAAGGCTGT 2040
ATTGTCTTCC TGTACCATTG CATTTCCTCA TTCCCAATTT GCACAAGGAT GTCTGGGTAA 2100
ACTATTCAAG AAATGGCTTT GAAATACAGC ATGGGAGCTT GTCTGAGTTG GAATGCAGAG 2160
TTGCACTGCA AAATGTCAGG AAATGGATGT CTCTCAGAAT GCCCAACTCC AAAGGATTTT 2220
ATATGTGTAT ATAGTAAGCA GTTTCCTGAT TCCAGCAGGC CAAAGAGTCT GCTGAATGTT 2280
GTGTTGCCGG AGACCTGTAT TTCTCAACAA GGTAAGATGG TATCCTAGCA ACTGCGGATT 2340
TTAATACATT TTCAGCAGAA GTACTTAGTT AATCTCTACC TTTAGGGATC GTTTCATCAT 2400
TTTTAGATGT TATACTTGAA ATACTGCATA ACTTTTAGCT TTCATGGGTT CCTTTTTTTC 2460
AGCCTTTAGG AGACTGTTAA GCAATTTGCT GTCCAACTTT TGTGTTGGTC TTAAACTGCA 2520
ATAGTAGTTT ACCTTGTATT GAAGAAATAA AGACCATTTT TATATTAAAA AATACTTTTG 2580
TCTGTCTTCA TTTTGACTTG TCTGATATCC TTGCAGTGCC CATTATGTCA GTTCTGTCAG 2640
ATATTCAGAC ATCAAAACTT AACGTGAGCT CAGTGGAGTT ACAGCTGCGG TTTTGATGCT 2700
```

(57) Abstract: The present invention relates to novel methods of producing transgenic avians, preferably chickens, wherein the incorporated transgene may be expressed as a constituent protein of the white of a hard-shell egg. The present invention provides sperm-mediated transfer for the introduction to an avian egg of a transgene encoding a heterologous polypeptide. The avian sperm may be irradiated before the transgenic gene is incorporated therein. Transgenic genes may be incorporated into avian sperm by lipofection, electroporation, restriction enzyme mediated integration (REMI) or similar methods. The modified avian sperm may then be delivered to an avian oocyte by microinjection, intracytoplasmic sperm injection (ICSI) or artificial insemination, or by natural coitus after the modified avian sperm are returned to a male bird. Heterologous nucleic acid may be integrated directly into the genomic nucleic acid of the oocyte or after first integrating the heterologous nucleic acid into the nucleic acid of a male germ cell and subsequent delivery of the transgenic male germ cell to an oocyte. Alternatively, the heterologous nucleic acid may be a episome within the sperm, or within the derivative zygote formed by the fusion of the sperm and the recipient oocyte, and may replicate independently of the zygote genome. Co-segregation of the episome with the replicated ooctye genome into all of the daughter cells may be induced by the heterologous nucleic acid having a centromeric body derived from, for example, a chromosome of a chicken.

WO 03/024199 A2

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

PRODUCTION OF TRANSGENIC AVIANS USING SPERM-MEDIATED TRANSFECTION

This application claims the benefit of United States Provisional Patent Application Serial No. 60/323,961, filed September 21, 2001, and United States Provisional Patent Application Serial No. 60/324,001, filed September 21, 2001, both of which are incorporated by reference herein in their entireties.

1. FIELD OF THE INVENTION

The present invention relates to methods of producing a transgenic avian by introducing a nucleic acid encoding a heterologous protein into the genome of an avian oocyte by sperm-mediated transfection. The present invention further relates generally to a transgenic avian capable of expressing a heterologous polypeptide, which, preferably is deposited into the white of an avian egg, said avian generated by sperm-mediated transgenesis. The invention further provides vectors containing coding sequences for heterologous proteins, the expression of which is under the control of a promoter and other regulatory elements that cause expression of the heterologous protein and preferably, lead to deposition of the protein in the avian egg. Also included in the invention are avian eggs derived from the transgenic avians and the heterologous proteins isolated therefrom.

2. BACKGROUND

The field of transgenics was initially developed to understand the action of a single gene in the context of the whole animal and the phenomena of gene activation, expression, and interaction. The technology has also been used to produce models for various diseases in humans and other animals and is amongst the most powerful tools available for the study of genetics and the understanding of genetic mechanisms and function. From an economic perspective, however, the use of transgenic technology for the production of specific proteins or other substances of pharmaceutical interest offers significant advantages over more conventional methods of protein production by gene expression. (Gordon *et al.*, 1987, *Biotechnology* 5: 1183-1187; Wilmut *et al.*, 1990, *Theriogenology* 33: 113-123).

In particular, the production of monoclonal antibodies by traditional methods is labor-intensive and costly. The purification of monoclonal antibodies from serum is a slow and low-yielding process. The use of hybridomas cell lines which were developed by fusing a B lymphocyte with a myeloma cell to propagate indefinitely *in vivo* as ascites, or *in vitro* in tissue culture requires major expenditures in tissue culture facilities or mice breeding.

(Kohler and Milstein, 1975, *Nature* 256: 495-497). Although various strategies have been proposed to overcome the deficiencies in antibody yield (*e.g.*, engineering single-chain antibodies (scAb) comprising immunoglobulin heavy and light chain variable regions), no method has been proven entirely satisfactory in elevating antibody yields to the levels desired for adequate commercial production.

The industry has been experimenting with transgenic animals that can express, for example, an exogenous protein such as an antibody under conditions that offer high yield of the protein in an active form while incorporating post-translational modifications, such as glycosylation, typically required for full functionality of the antibody. In this context, heterologous nucleic acids have been engineered so that an expressed protein may be joined to a protein or peptide that will allow secretion of the transgenic expression product into milk or urine, from which the protein may then be recovered. These procedures have had limited success, however, and may require lactating animals, with the attendant costs of maintaining individual animals or herds of large species, including cows, sheep or goats.

Avian Transgenics

One transgenic system that holds potential is the avian reproductive system. The exogenous protein can be produced in the white of an avian egg from which it may be readily purified. (MacArthur, PCT Publication WO 97/47739). The production of an avian egg begins with formation of a large yolk in the ovary of the hen. The unfertilized oocyte or ovum is positioned on top of the yolk sac. After ovulation, the ovum passes into the infundibulum of the oviduct where it is fertilized, if sperm are present, and then moves into the magnum of the oviduct, lined with tubular gland cells. These cells secrete the egg-white proteins, including ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin, into the lumen of the magnum where they are deposited onto the avian embryo and yolk.

The hen oviduct, for example, can serve as an excellent protein bioreactor because of the high levels of protein production, the promise of proper folding and post-translation modification of the target protein, the ease of product recovery, and the shorter developmental period of chickens compared to other potential animal species. The economic advantage of breeding flocks of transgenic birds laying eggs expressing exogenous proteins would be significant when compared to more traditional animals, such as cows, sheep or goats, producing heterologous protein in milk. What is needed, however, is an efficient method of introducing a heterologous nucleic acid into a recipient avian embryonic cell.

Vectors

Genetic information has been transferred to avian embryos using vectors. Bosselman *et al.* in U.S. Patent No. 5,162,215 describes a method for introducing a replication-defective retroviral vector into a pluripotent stem cell of an unincubated chick embryo, and further describes chimeric chickens whose cells express a heterologous vector nucleic acid sequence. However, the percentage of G1 transgenic offspring (progeny from vector-positive male G0 birds) was low and varied between 1% and approximately 8%. In addition, the use of viral vectors poses limitations, including limitations on transgene size and potential viral infection of the offspring, thus, posing significant regulatory issues for production of biologics.

Similarly, Jaenisch reported that while retroviral vectors did transfer genetic information to embryos, the resulting animals were mosaics with gene insertions at various loci in the genomic nucleic acid. (1976, *Proc. Natl. Acad. Sci. USA* 73: 1260-1264). The transgenes were also differentially expressed in the different tissues of each animal. (Jaenisch, 1980, *Cell* 19: 181-188).

Nuclear Transfer

Nuclear transfer from cultured cell populations is another route to produce transgenics, wherein donor cells may be sexed, optionally genetically modified, and then selected in culture before their use. The resultant transgenic animal originates from a single transgenic nucleus and therefore, mosaics are avoided. Nuclear transfer from cultured somatic cells also provides a route for directed genetic manipulation of animal species, including the addition or "knock-in" of genes, and the removal or inactivation or "knock-out" of genes or their associated control sequences (Polejaeva *et al.*, 2000, *Theriogenology* 53: 117-26).

Two types of recipient cells are commonly used in nuclear transfer procedures: oocytes arrested at the metaphase of the second meiotic division (MII) and which have a metaphase plate with the chromosomes arranged on the meiotic spindle, and pronuclear zygotes. In agricultural mammals, however, development does not always occur when pronuclear zygotes are used, and, therefore, MII-arrested oocytes are the preferred recipient cells. Enucleated two-cell stage blastomeres of mice have also been used as recipients.

After enucleation and introduction of donor genetic material, the reconstructed embryo is cultured to the morula or blastocyte stage, and transferred to a recipient animal, either *in vitro* or *in vivo*, and developed to term. (Eyestone and Campbell, 1999, *J. Reprod. Fertil. Suppl.* 54: 489-97). Double nuclear transfer has been reported in which an activated,

previously transferred nucleus is removed from the host unfertilized egg and transferred again into an enucleated fertilized embryo. Activation (initiation of development) is most often induced chemically. Cultured cells can also be frozen and stored indefinitely for future use.

Although gene targeting techniques combined with nuclear transfer hold tremendous promise for nutritional and medical applications, current approaches suffer from several limitations, including long generation times between the founder animal and production transgenic herds, and extensive husbandry and veterinary costs. It is therefore desirable to use a system where cultured somatic cells for nuclear transfer are more efficiently employed.

Sperm-Mediated Transfection Mechanism

A promising method for producing transgenic animals is the stable transfection of male germ cells *in vitro* and their transfer to a recipient oocyte. PCT Publication WO 87/05325 discloses a method of transferring organic and/or inorganic material into sperm or egg cells by using liposomes. Bachiller *et al.* used Lipofectin-based liposomes to transfer DNA into mice sperm, and provided evidence that the liposome transfected DNA was overwhelmingly contained within the sperm's nucleus. (1991, *Mol. Reprod. Develop.* 30: 194-200). However, no transgenic mice could be produced by this technique.

Similarly, Nakanishi and Iritani used Lipofectin-based liposomes to associate heterologous DNA with chicken sperm, which were in turn used to artificially inseminate hens. (1993, *Mol. Reprod. Develop.* 36: 258-261). Although the heterologous DNA was detectable in many of the resultant fertilized eggs, there was no evidence of genomic integration of the heterologous DNA either in the DNA-liposome treated sperm or in the resultant chicks.

Heterologous DNA may also be transferred into sperm cells by a process called electroporation that creates temporary, short-lived pores in the cell membrane of living cells by exposing them to a sequence of brief electrical pulses of high field strength. The pores allow cells to take up heterologous material such as DNA, while only slightly compromising cell viability. Gagne *et al.* discloses the use of electroporation to introduce heterologous DNA into bovine sperm subsequently used to fertilize ova. (1991, *Mol. Reprod. Develop.* 29: 6-15). However, there was no evidence of integration of the electroporated DNA either in the sperm nucleus or in the nucleus of the egg subsequent to fertilization by the sperm.

Yet another method initially developed for integrating heterologous DNA into yeasts and slime molds, and later adapted to avian sperm, is restriction enzyme mediated

integration (REMI), which utilizes a linear DNA derived from a plasmid DNA by cutting that plasmid with a restriction enzyme that generates single-stranded cohesive ends. (Shemesh *et al.*, PCT International Publication WO 99/42569). The linear, cohesive-ended DNA together with the restriction enzyme used to produce the cohesive ends is then introduced into the target cells by electroporation or liposome transfection. The restriction enzyme is then thought to cut the genomic DNA at sites that enable the heterologous DNA to integrate via its matching cohesive ends. (Schiestl and Petes, 1991, *Proc. Natl. Acad. Sci. USA* 88: 7585-7589). Although Shemesh described transgenic birds that were resistant to Infectious Bursal Disease, there was no evidence of expression or deposition of a heterologous protein in the oviduct for deposition onto egg whites.

What is needed, therefore, is an efficient method of generating a transgenic avian capable of expressing a heterologous protein coded by a transgene, particularly in the oviduct for deposition into egg whites.

3. SUMMARY OF THE INVENTION

The invention provides methods for the stable introduction by sperm-mediated transfection of heterologous coding sequences into the genome of an avian, preferably a chicken, and expressing those heterologous coding sequence to produce desired proteins and/or to alter the phenotype of the transgenic avian. Synthetic vectors and gene promoters useful in the methods are also provided by the present invention, as are transgenic avians that express a heterologous protein and avian eggs, preferably chicken eggs, containing a heterologous protein. In a preferred embodiment, the vectors useful in methods of the invention are not eukaryotic viral, more preferably not retroviral, vectors (although the vectors may contain transcriptional regulatory elements, such as promoters, from eukaryotic viruses). In other embodiments, however, the vectors are retroviral vectors.

One aspect of the present invention is a method of producing a transgenic avian, preferably a chicken, by introducing in an avian oocyte at least one transgene encoding at least one heterologous polypeptide by sperm-mediated transfection. The method comprises first, isolating an avian sperm, second, incorporating a transgene into the avian sperm, and third, delivering the modified avian sperm to an avian oocyte. In one embodiment, the avian sperm is irradiated with gamma rays before the transgene is incorporated therein.

In one embodiment, the transgene is injected directly into the testis of a male avian and incorporated in the avian sperm. The modified sperm is then delivered to the avian oocyte by mating the male avian with a wild type or transgenic female avian.

In another embodiment, the transgene is incorporated in the avian sperm *in vitro* by lipofection, electroporation, restriction enzyme mediated integration (REMI) or similar methods. In a preferred embodiment, the modified avian sperm is then delivered to the avian oocyte by natural coitus after the modified avian sperm are returned to the testis of a male avian. In another preferred embodiment, the modified avian sperm is delivered to the avian oocyte by microinjection (*e.g.*, intracytoplasmic sperm injection (ICIS) or standard artificial fertilization methods). The resulting transgenic embryo can then be transferred to the oviduct of a recipient hen for development and to be laid as a shelled egg (or, alternatively, cultured *ex vivo*). The shelled egg is incubated to hatch a transgenic avian that has incorporated, preferably integrated into its genome, the selected nucleic acid. In preferred embodiments, the avian sperm is first irradiated before incorporated with the transgene.

In certain embodiments, a transgene comprising a heterologous nucleic acid may be integrated directly into the genomic nucleic acid of an avian sperm and subsequently delivered to an avian oocyte. When the heterologous nucleic acid is directly integrated into the genome of the avian sperm which then fertilizes an avian oocyte, the resulting transgenic embryo will include the transgenic heterologous nucleic acid in all of its cells. In preferred embodiments, the transgenic heterologous nucleic acid is incorporated into at least one embryonic cell, preferably the germinal disk of an early stage embryo, that then develop into a transgenic avian.

Alternatively, the heterologous nucleic acid may be an episome within the modified avian sperm, or within the derivative zygote formed by the fusion of the modified avian sperm and the avian oocyte. The episome may replicate independently of the zygote genome. When the heterologous nucleic acid is episomal with respect to the genome of the transgenic zygote, and the episomal nucleic acid has a centromeric body, most, if not all, of the cells of the transgenic embryo will include the heterologous nucleic acid. Accordingly, in preferred embodiments, the transgene further comprises centromere and/or telomere sequences of an avian chromosome.

The invention further provides method for incorporating at least one transgene into the genome of a spermatozoon cell or a precursor thereof isolated from a donor male avian, and returning the modified cell to the testis of a recipient male avian, preferably the donor male avian, so that a genetically modified male gamete is produced by the male avian. Breeding the male avian with a female of its species will generate a transgenic progeny carrying the at least one transgene in its genome.

The invention also provides methods for introducing a heterologous nucleic acid to an avian oocyte in addition to those described in United States Application Serial No. 09/877,374, filed June 8, 2001, entitled "Production of Monoclonal Antibody By a Transgenic Chicken", by Jeffrey C. Rapp; and United States Application Serial No. _____, filed September 18, 2002, entitled "Production of a Transgenic Avian By Cytoplasmic Injection", by Jeffrey C. Rapp and Leandro Christmann, both of which are incorporated by reference herein in their entireties. In certain embodiments, the avian oocyte is removed from the ovaries of a donor female avian to facilitate *in vitro* fertilization by the modified avian sperm of the invention. In other certain embodiments, the modified avian sperm is delivered to an avian oocyte *in vivo* by natural coitus. The fertilized ova is then, preferably, returned to or maintained in the oviduct of the donor female avian or a surrogate female avian to be laid as a hard-shell egg or, as an alternative, cultured *ex vivo*. The hard-shell egg is incubated and hatched, producing a transgenic chick that expresses a heterologous protein and/or that can be bred to generate a line of transgenic avians expressing a heterologous protein.

Preferably, the avian sperm or the reproductive system of a male avian, preferably the seminiferous tubules and/or site of sperm production, development, and/or storage in the testis, is irradiated by gamma rays before transgene incorporation. More preferably, the transgene is integrated directly into the genome of the avian sperm. Most preferably, the transgene further comprises centromere and/or telomere sequences.

In particular embodiments, the level of mosaicism of the transgene (percentage of cells containing the transgene) in avians hatched from sperm-mediated transfected embryos (*i.e.*, the G0s) is greater than 5%, 10%, 25%, 50%, 75% or 90%, or is the equivalent of one copy per one genome, two genomes, five genomes, seven genomes or eight genomes, as determined by any number of techniques known in the art and described *infra*. In additional particular embodiments, the percentage of G0s that transmit the transgene to progeny (G1s) is greater than 5%, preferably, greater than 10%, 20%, 30%, 40%, and, most preferably, greater than 50%.

In certain other embodiments, the level of transgenics that result from mating with a wild type or transgenic avian avians hatched from sperm-mediated transfected embryos (*i.e.*, the G0s) is greater than 5%, 10%, 25%, 50%, 75% or 90%.

In another embodiment, the present invention provides methods for producing heterologous proteins in avians. Transgenes are introduced by sperm-mediated transfection into the genome of an avian oocyte which becomes fertilized and then develops into a transgenic avian. The heterologous protein(s) of interest may be expressed in the tubular

gland cells of the magnum of the oviduct, secreted into the lumen, and most preferably, deposited within the egg white onto the egg yolk or expressed, for example, in the serum of the avian. In preferred embodiments, the level of expression of the heterologous protein in the egg white of eggs laid by G0 and/or G1 chicks and/or their progeny is greater than 5 μg, 10 μg, 50 μg, 100 μg, 250 μg, 500 μg or 750 μg, more preferably greater than 1 mg, 2 mg, 5 mg, 10 mg, 20 mg, 50 mg, 100 mg, 200 mg, 500 mg, 700 mg, 1 gram, 2 grams, 3 grams, 4 grams or 5 grams.

The transgenic avians can also be bred to identify those avians that carry the transgene in their germ line. The exogenous gene coding for the heterologous proteins can therefore be transmitted by sperm-mediated transfection of the exogenous gene into the avian oocytes, and by subsequent stable transmission of the exogenous gene to the avian's offspring in a Mendelian fashion. More information on Mendelian inheritance can be found in Hartl and Jones, 2001, Genetics: Analysis of Genes and Genomes, 5th ed., Jones & Bartlett Publishers, Inc., the content of which is incorporated by reference herein in its entirety.

Another aspect of the invention provides for the isolation of heterologous proteins in transgenic avians and the use thereof in pharmaceutical products including but not limited to vaccines, biologics and, particularly, therapeutically or diagnostically useful antibodies. The expressed heterologous protein(s) of interest may be collected and processed using standard techniques from the avian eggs, preferably the egg white, the serum, or other tissues from the transgenic avian.

The present invention further provides methods for producing a heterologous protein in an avian oviduct. The method comprises, as a first step, providing a vector containing a coding sequence and a promoter that functions in avians, preferably in the avian magnum, operably linked to the coding sequence, so that the promoter can effect expression of the nucleic acid in the tubular gland cells of the magnum of an avian oviduct and/or in any other desired tissue of the avian. In a preferred embodiment, the vector containing the transgene is not a eukaryotic viral vector (preferably, not a retroviral vector, such as but not limited to reticuloendotheliosis virus (REV), ALV or MMLV) or derived from a eukaryotic virus (but, in certain embodiments, may contain promoter and/or other gene expression regulatory sequences from a eukaryotic virus, such as, but not limited to, a cytomegalovirus promoter). Next, the vector is introduced into avian sperm *in vitro* by lipofection, electroporation, restriction enzyme mediated integration (REMI) or similar methods, or *in vivo* by directly injecting into the testis, so that the vector sequence may be incorporated into the avian sperm. In preferred embodiments, the avian sperm or precursor cells are irradiated by

gamma rays before the vector sequence is incorporated therein. In another preferred embodiment, the vector sequence further comprises centromere and/or telomere sequence. Then, the modified avian sperm are delivered to an avian oocyte by natural coitus or *in vitro* by microinjection or artificial insemination to form a transgenic embryonic cell. In certain embodiments, the recipient avian oocyte is wild type unmodified or preferably, modified in a manner that facilitates the delivery of transgene by the modified avian sperm. In certain other embodiments, the recipient avian oocyte is derived from a first-generation or preferably, second-generation transgenic avian whose germ-line carries the transgene. Finally, a mature transgenic avian that expresses the exogenous protein in its oviduct is derived from the transgenic embryonic cell or by breeding a transgenic avian derived from the transgenic embryonic cell.

The present invention further provides promoters useful for expression of the heterologous protein in the egg. For example, the transgene may comprise regions of at least two promoters derived from an avian including, but not limited to, an oviduct-specific promoter such as ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin promoter or any other promoter that directs expression of a gene in an avian, particularly in a specific tissue of interest, such as the magnum, and a protamine promoter, or a fragment thereof which is sufficient to drive the expression of a marker gene such as Green Fluorescent Protein (GFP). Alternatively, the promoter used in the expression vector may be derived from that of the *lysozyme* gene that is expressed in both the oviduct and macrophages. In particular embodiments, the gene regulatory sequences are flanked by matrix attachment regions (MARs), preferably, but not limited to those associated with the *lysozyme* gene in chickens or other avians. The nucleic acid encoding the polypeptide may be operably linked to a transcription promoter and/or a transcription terminator.

Other embodiments of the invention provide for transgenic avians, such as chickens or quail, carrying a transgene in the genetic material of their germ-line tissue, preferably where the transgene was not introduced into the avian genome using a eukaryotic viral promoter. The transgene incorporated into the genomic DNA of a recipient avian can encode at least one polypeptide that may be, for example, but is not limited to, a cytokine, a growth factor, enzyme, structural protein, immunoglobulin, or any other polypeptide of interest that is capable of being expressed by an avian cell or tissue. Preferably, the heterologous protein is a mammalian, preferably a human, protein or derived from a mammalian, or preferably a human, protein (*e.g.*, a derivative or variant thereof). In particular embodiments, the invention provides heterologous proteins isolated or purified from an avian tissue, preferably serum, more preferably eggs, most preferably egg whites,

and pharmaceutical compositions comprising such heterologous proteins. In a more preferred embodiment, the heterologous protein is an antibody that is human (including antibodies produced from human immunoglobulin sequences in mice or in antibody libraries or synthetically produced but having variable domain framework regions that are the same as or homologous to human framework regions) or humanized.

The present invention further relates to nucleic acid vectors (preferably, not derived from eukaryotic viruses, except, in certain embodiments, for eukaryotic viral promoters and/or enhancers) and transgenes inserted therein that incorporate multiple polypeptide-encoding regions, wherein a first polypeptide-encoding region is operatively linked to a transcription promoter and a second polypeptide-encoding region is operatively linked to an Internal Ribosome Entry Sequence (IRES). For example, the vector may contain coding sequences for two different heterologous proteins (*e.g.*, the heavy and light chains of an immunoglobulin) or the coding sequences for all or a significant part of the genomic sequence for the gene from which the promoter driving expression of the transgene is derived, and the heterologous protein desired to be expressed (*e.g.*, a construct containing the genomic coding sequences, including introns, of the avian *lysozyme* gene when the avian lysozyme promoter is used to drive expression of the transgene, an IRES, and the coding sequence for the heterologous protein desired to be expressed downstream (*i.e.*, 3' on the RNA transcript of the IRES). Thus, in certain embodiments, the nucleic acid encoding the heterologous protein is introduced into the 5' untranslated or 3' untranslated regions of an endogenous gene, such as but not limited to, ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin, with an IRES sequence directing translation of the heterologous sequence.

Such nucleic acid constructs, when inserted into the genome of an avian and expressed therein, will generate individual polypeptides that may be post-translationally modified, for example, glycosylated or, in certain embodiments, form complexes, such as heterodimers with each other in the white of the avian egg. Alternatively, the expressed polypeptides may be isolated from an avian egg and combined *in vitro*, or expressed in a non-reproductive tissue such as serum. In other embodiments, for example, but not limited to, when expression of both heavy and light chains of an antibody is desired, two separate constructs, each containing a coding sequence for one of the heterologous proteins operably linked to a promoter (either the same or different promoters), are introduced into embryonic cells by sperm-mediated transfection to generate transgenic avians that harbor both transgenes in their genomes and expressing both heterologous proteins are identified. Alternatively, two transgenic avians each containing one of the two heterologous proteins

(*e.g.*, one transgenic avian having a transgene encoding the light chain of an antibody and a second transgenic avian having a transgene encoding the heavy chain of the antibody) can be bred by Mendelian genetics to obtain an avian containing both transgenes in its germline and expressing both transgene encoded proteins, preferably in eggs. (See Hartl and Jones, 2001, Genetics: Analysis of Genes and Genomes, 5th ed., Jones & Bartlett Publishers, Inc., the content of which is incorporated by reference herein in its entirety).

For convenience, certain terms employed in the specification, examples, and appended claims are collected here.

Additional objects and aspects of the present invention will become more apparent upon review of the detailed description set forth below when taken in conjunction with the accompanying figures, which are briefly described as follows.

3.1 DEFINITIONS

The term "animal" as used herein refers to all vertebrate animals, including birds. It also includes an individual animal in all stages of development, including embryonic and fetal stages.

The term "avian" as used herein refers to any species, subspecies or race of organism of the taxonomic class *aves*, such as, but not limited to, chicken, quail, turkey, duck, goose, pheasants, parrots, finches, hawks, crows and ratites including ostrich, emu and cassowary. The term includes the various known strains of *Gallus gallus*, or chickens, (for example, White Leghorn, Brown Leghorn, Barred-Rock, Sussex, New Hampshire, Rhode Island, Ausstralorp, Minorca, Amrox, California Gray, Italian Partridge-colored), as well as strains of turkeys, pheasants, quails, duck, ostriches and other poultry commonly bred in commercial quantities.

The term "male germ cells" as used herein refers to sperm, sperm cells, spermatozoa (*i.e.*, male gametes) and developmental precursors thereof. Male germ cells with the capacity to swim and transfer nucleic acid to an ovum are herein referred to as "viable male germ cells." In fetal development, primordial germ cells are thought to arise from the embryonic ectoderm, and are first seen in the epithelium of the endodermal yolk sac at the E8 stage. From there they migrate through the hindgut endoderm to the genital ridges. In the sexually mature male vertebrate animal, there are several types of cells that are precursors of spermatozoa, and which can be genetically modified, including the primitive spermatogonial stem cells, known as A0/As, which differentiate into type B spermatogonia. The latter further differentiate to form primary spermatocytes, and enter a prolonged meiotic prophase during which homologous chromosomes pair and recombine. Useful precursor

cells at several morphological/developmental stages are also distinguishable: preleptotene spermatocytes, leptotene spermatocytes, zygotene spermatocytes, pachytene spermatocytes, secondary, spermatocytes, and the haploid spermatids. The latter undergo further morphological changes during spermatogenesis, including the reshaping of their nucleus, the formation of aerosome, and assembly of the tail. The final changes in the spermatozoon (i.e., male gamete) take place in the genital tract of the female, prior to fertilization.

The terms "ovum" and "oocyte" are used interchangeably herein. Although only one ovum matures at a time, an animal is born with a finite number of ova. In avian species, such as a chicken, ovulation, which is the shedding of an egg from the ovarian follicle, occurs when the brain's pituitary gland releases a luteinizing hormone, LH. Mature follicles form a stalk or pedicel of connective tissue and smooth muscle. Immediately after ovulation the follicle becomes a thin-walled sac, the post-ovulatory follicle. The mature ovum erupts from its sac and starts its journey through the oviduct. Eventually, the ovum enters the infundibulum where fertilization occurs. Fertilization must take place within 15 minutes of ovulation, before the ovum becomes covered by albumen. During fertilization, sperm (avians have polyspermic fertilization) penetrate the blastodisc. When the sperm lodges within this germinal disk, an embryo begins to form as a "blastoderm" or "zygote."

The term "embryonic cells" as used herein refers to cells that are typically single cell embryos, fertilized or unfertilized, or the equivalent thereof, and is meant to encompass dividing embryos, such as two-cell, four-cell, or even later stages as described by Eyal-Giladi and Kochav (1976, *Dev. Biol.* 49: 321-337) and ova 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, or 20 hours after the preceding lay. The embryonic cells may be isolated freshly, maintained in culture, or reside within an embryo.

The term "fragment" as used herein to refers to an at least 10, 20, 50, 75, 100, 150, 200, 250, 300, 500, 1000, 2000 or 5000 nucleotide long portion of a nucleic acid (*e.g.*, cDNA) that has been constructed artificially (*e.g.*, by chemical synthesis) or by cleaving a natural product into multiple pieces, using restriction endonucleases or mechanical shearing, or enzymatically, for example, by PCR or any other polymerizing technique known in the art, or expressed in a host cell by recombinant nucleic acid technology known to one of skill in the art. The term "fragment" as used herein may also refer to an at least 5, 10, 20, 30, 40, 50, 75, 100, 150, 200, 250, 300, 400, 500, 1000, 2000 or 5000 amino acid portion of a polypeptide, which portion is cleaved from a naturally occurring polypeptide by proteolytic cleavage by at least one protease, or is a portion of the naturally occurring polypeptide synthesized by chemical methods or using recombinant DNA technology (*e.g.*, expressed

from a portion of the nucleotide sequence encoding the naturally occurring polypeptide) known to one of skill in the art.

The term "isolated nucleic acid" as used herein refers to a nucleic acid that has been removed from other components of the cell containing the nucleic acid or from other components of chemical/synthetic reaction used to generate the nucleic acid. In specific embodiments, the nucleic acid is 50%, 60%, 70%, 80%, 90%, 95%, 99% or 100% pure. The "isolated nucleic acid" is neither (a) identical to that of any naturally occurring nucleic acid nor (b) identical to that of any fragment of a naturally occurring genomic nucleic acid spanning more than three separate genes, and includes DNA, RNA, or derivatives or variants thereof. The term covers, for example, (a) a DNA which has the sequence of part of a naturally occurring genomic molecule but is not flanked by at least one of the coding sequences that flank that part of the molecule in the genome of the species in which it naturally occurs; (b) a nucleic acid incorporated into a vector or into the genomic nucleic acid of a prokaryote or eukaryote in a manner such that the resulting molecule is not identical to any vector or naturally occurring genomic DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), ligase chain reaction (LCR) or chemical synthesis, or a restriction fragment; (d) a recombinant nucleotide sequence that is part of a hybrid gene, *i.e.*, a gene encoding a fusion protein; and (e) a recombinant nucleotide sequence that is part of a hybrid sequence that is not naturally occurring. The techniques used to isolate and characterize the nucleic acids and proteins of the present invention are well known to those of skill in the art and standard molecular biology and biochemical manuals may be consulted to select suitable protocols without undue experimentation. See, *e.g.*, Sambrook *et al*, Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor Press (2001); the content of which is herein incorporated by reference in its entirety.

By the use of the term "enriched" in reference to nucleic acid it is meant that the specific DNA or RNA sequence constitutes a significantly higher fraction of the total DNA or RNA present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. Enriched does not imply that there are no other DNA or RNA sequences present, just that the relative amount of the sequence of interest has been significantly increased, for example, by 1 fold, 2 fold, 5 fold, 10 fold, 50 fold, 100 fold, 500 fold, 1000 fold, 10,000 fold, 100,000 fold or 1,000,000 fold. The other DNA may, for example, be derived from a yeast or bacterial genome, or a cloning vector, such as a plasmid or a viral vector.

The term "transcription regulatory sequences" as used herein refers to nucleotide sequences that are associated with a gene nucleic acid sequence and that regulate the transcriptional expression of the gene. The "transcription regulatory sequences" may be isolated and incorporated into a vector nucleic acid to enable regulated transcription in appropriate cells of portions of the vector DNA. Exemplary transcription regulatory sequences include enhancer elements, hormone response elements, steroid response elements, negative regulatory elements, and the like. The "transcription regulatory sequences" may be isolated and incorporated into a vector nucleic acid to enable regulated transcription in appropriate cells of portions of the vector DNA. The "transcription regulatory sequence" may precede, but is not limited to, the region of a nucleic acid sequence that is in the region 5' of the end of a protein coding sequence that may be transcribed into mRNA. Transcriptional regulatory sequences may also be located within a protein coding region, in regions of a gene that are identified as "intron" regions, or may be in regions of nucleic acid sequence that are in the region of nucleic acid.

The term "promoter" as used herein refers to the DNA sequence that determines the site of transcription initiation by an RNA polymerase. A "promoter-proximal element" may be a regulatory sequence within about 200 base pairs of the transcription start site. A "magnum-specific" promoter, as used herein, is a promoter that is primarily or exclusively active in the tubular gland cells of the avian magnum. Useful promoters also include exogenously inducible promoters. These are promoters that can be "turned on" in response to an exogenously supplied agent or stimulus, which is generally not an endogenous metabolite or cytokine. Examples include an antibiotic-inducible promoter, such as a tetracycline-inducible promoter, a heat-inducible promoter, a light-inducible promoter, or a laser inducible promoter. (See, *e.g.*, Halloran *et al.*, 2000, *Development* 127: 1953-1960; Gemer *et al.*, 2000, *Int. J. Hyperthermia* 16: 171-81; Rang and Will, 2000, *Nucleic Acids Res.* 28: 1120-5; Hagihara *et al.*, 1999, *Cell Transplant* 8: 4314; Huang *et al.*, 1999, *Mol. Med.* 5: 129-37; Forster *et al.*, 1999, *Nucleic Acids Res.* 27: 708-10; Liu *et al.*, 1998, *Biotechniques* 24: 624-8, 630-2; the contents of which have been incorporated herein by reference in their entireties).

To facilitate manipulation and handling of the nucleic acid to be administered, the nucleic acid is preferably inserted into a cassette where it is operably linked to a promoter. The promoter should be capable of driving expression in the desired cells. The selection of appropriate promoters can be readily accomplished. For some applications, a high expression promoter is preferred such as the cytomegalovirus (CMV) promoter. Other promoters useful in the present invention include the Rous Sarcoma Virus (RSV) promoter

(Davis *et al.*, 1993, *Hum. Gene Therap.* 4:151). In other embodiments, all or a portion of the, for example, lysozyme, ovomucoid, albumin, conalbumin or ovotransferrin promoters, which direct expression of proteins present in egg white, are used, as detailed *infra*, or synthetic promoters such as the MDOT promoter described *infra*.

The term "expressed" or "expression" as used herein refers to the transcription from a gene to give an RNA nucleic acid molecule complementary at least in part to a region of one of the two nucleic acid strands of the gene. The term "expressed" or "expression" as used herein also refers to the translation from said RNA nucleic acid molecule to give a protein or polypeptide or a portion thereof.

The term "matrix attachment regions" as used herein refers to DNA sequences having an affinity or intrinsic binding ability for the nuclear scaffold or matrix. The MAR elements of the chicken lysozyme locus were described by Phi-Van *et al.*, 1996, *E.M.B.O. J.* 76:665-664 and Phi-Van, L. and Stratling, W.H., 1996, *Biochem.* 35:10735-10742; incorporated herein by reference in their entireties.

The term "nucleic acid vector" as used herein refers to a natural or synthetic single or double stranded plasmid or viral nucleic acid molecule, or any other nucleic acid molecule, such as but not limited to YACs, BACs, bacteriophage-derived artificial chromosome (BBPAC), cosmid or P1 derived artificial chromosome (PAC), that can be transfected or transformed into cells and replicate independently of, or within, the host cell genome. A circular double stranded vector can be linearized by treatment with an appropriate restriction enzyme based on the nucleotide sequence of the vector. A nucleic acid can be inserted into a vector by cutting the vector with restriction enzymes and ligating the pieces together. The nucleic acid molecule can be RNA or DNA.

The term "expression vector" as used herein refers to a nucleic acid vector that comprises regulatory sequences operably linked to a nucleotide sequence coding for at least one polypeptide. As used herein, the term "regulatory sequences" includes promoters, enhancers, and other elements that may control expression. Standard molecular biology textbooks such as Sambrook *et al.*, (*supra*) and Lodish *et al.*, eds "Molecular Cell Biology" Freeman (2000) and incorporated herein by reference in their entireties, may be consulted to design suitable expression vectors, promoters, and other expression control elements. It should be recognized, however, that the choice of a suitable expression vector depends upon multiple factors including the choice of the host cell to be transformed and/or the type of protein to be expressed. Also useful for various applications are tissue-selective (*i.e.*, tissue-specific) promoters, *i.e.*, promoters from which expression occurs preferentially in cells or a particular kind of tissue, compared to one or more other types of tissue. For example,

chicken oviduct-specific promoters naturally associated with the proteins of avian egg whites including, but not limited to, lysozyme, ovomucoid, albumin, conalbumin, and ovotransferrin may be used.

The term "recombinant cell" refers to a cell that has a new combination of nucleic acid segments that are not covalently linked to each other in nature in that particular configuration. A new combination of nucleic acid segments can be introduced into an organism using a wide array of nucleic acid manipulation techniques available to those skilled in the art. A recombinant cell can be a single eukaryotic cell, or a single prokaryotic cell, or a mammalian cell. The recombinant cell may harbor a vector that is extragenomic. An extragenomic nucleic acid vector does not insert into the cell's genome. A recombinant cell can further harbor a vector or a portion thereof (*e.g.*, the portion containing the regulatory sequences and the coding sequence) that is intragenomic. The term "intragenomic" defines a nucleic acid construct incorporated within the recombinant cell's genome.

The terms "recombinant nucleic acid" and "recombinant DNA" as used herein refer to a combination of at least two nucleic acid sequences that is not naturally found in a eukaryotic or prokaryotic cell in that particular configuration. The nucleic acid sequences may include, but are not limited to, nucleic acid vectors, gene expression regulatory elements, origins of replication, sequences that when expressed confer antibiotic resistance, and protein-encoding sequences. The term "recombinant polypeptide" is meant to include a polypeptide produced by recombinant DNA techniques such that it is distinct from a naturally occurring polypeptide either in its location, purity or structure. Generally, such a recombinant polypeptide will be present in a cell in an amount different from that normally observed in nature.

As used herein, the term "transgene" refers to a nucleic acid sequence (encoding, for example, a human interferon polypeptide) that is partly or entirely heterologous, *i.e.*, foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (*e.g.*, it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene also includes a regulatory sequence designed to be inserted into the genome such that it regulates the expression of an endogenous coding sequence, *e.g.*, to increase expression and or to change the timing and or tissue specificity of expression, etc. (*e.g.*, to effect "gene activation").

The term "transgenic animal" as used herein refers to an animal, including an avian species such as a chicken, in which one or more of the cells of the animal contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art or by the methods of the present invention.

As used herein, a "transgenic avian" is any avian species, including but not limited to, chicken, turkey, duck, goose, quail, pheasants, parrots, finches, hawks, crows and ratites including ostrich, emu and cassowary, in which one or more of the cells of the avian may contain heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques known in the art, and particularly, as described herein. The nucleic acid is introduced into a cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization (although it does include fertilization with sperm into which a transgene has been introduced, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In the typical transgenic avian, the transgene causes cells to express a recombinant form of the subject polypeptide, *e.g.* either agonistic or antagonistic forms, or a form in which the gene has been disrupted.

The terms "chimeric animal" or "mosaic animal" are used herein to refer to animals in which the recombinant gene is found, or in which the recombinant is expressed in some but not all cells of the animal. The term "tissue-specific chimeric animal" indicates that the polypeptide encoding gene is present and expressed in some tissues, but not others.

The term "knock-in animal" refers to an animal that carries a specific nucleic acid sequence such as a "knock-in sequence" in a predetermined coding or noncoding region, wherein the knock-in sequence is introduced through methods of recombination, such as homologous recombination. The recombination event comprises replacing all or part of a gene of the animal by a functional homologous gene or gene segment of another animal, where the respective knock-in sequence is placed in the genomic sequence.

The term "chromosomal positional effect (CPE)" as used herein refers to the variation in the degree of gene transcription as a function of the location of the transcribed locus within the cell genome. Random transgenesis may result in a transgene being inserted at different locations in the genome so that individual cells of a population of transgenic cells may each have at least one transgene, each at a different location and therefore each in a different genetic environment. Each cell, therefore, may express the transgene at a level specific for that particular cell and dependant upon the immediate genetic environment of

the transgene. In a transgenic avian, as a consequence, different tissues may exhibit different levels of transgene expression.

The term "cytokine" as used herein refers to any secreted polypeptide that affects the functions of cells and is a molecule that modulates interactions between cells in the immune, inflammatory or hematopoietic response. A cytokine includes, but is not limited to, monokines and lymphokines regardless of which cells produce them. For instance, a monokine is generally referred to as being produced and secreted by a mononuclear cell, such as a macrophage and/or monocyte. Many other cells however also produce monokines, such as natural killer cells, fibroblasts, basophils, neutrophils, endothelial cells, brain astrocytes, bone marrow stromal cells, epidermal keratinocytes and B-lymphocytes. Lymphokines are generally referred to as being produced by lymphocyte cells. Examples of cytokines include, but are not limited to, Interleukin-1 (IL-1), Interleukin-6 (IL-6), Interleukin-8 (IL-8), Tumor Necrosis Factor-alpha (TNF-alpha) and Tumor Necrosis Factor β (TNF-β).

The term "antibody" as used herein refers to polyclonal and monoclonal antibodies and fragments thereof, and immunologic binding equivalents thereof. The term "antibody" refers to a homogeneous molecular entity, or a mixture such as a polyclonal serum product made up of a plurality of different molecular entities, and may further comprise any modified or derivatised variant thereof that retains the ability to specifically bind an epitope. A monoclonal antibody is capable of selectively binding to a target antigen or epitope. Antibodies may include, but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, disulfide-linked Fvs (sdFv) fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, intrabodies, synthetic antibodies, and epitope-binding fragments of any of the above.

The term "immunoglobulin polypeptide" as used herein refers to a polypeptide derived from a constituent polypeptide of an immunoglobulin. An "immunoglobulin polypeptide" may be, but is not limited to, an immunoglobulin (preferably an antibody) heavy or light chain and may include a variable region, a diversity region, joining region and a constant region or any combination, variant or truncated form thereof. The term "immunoglobulin polypeptide" further includes single-chain antibodies comprised of, but not limited to, an immunoglobulin heavy chain variable region, an immunoglobulin light chain variable region and optionally a peptide linker.

The term "origin of replication" (*ori*) as used herein refers to unique regions of a nucleic acid sequence containing multiple short repeated sequences, recognized by

multimeric origin of replication binding proteins that organize the assembly of multiple enzymes and proteins required for the replication of the nucleic acid. The origin of replication derived from *E. coli* may be included in a plasmid for replication of the plasmid in a bacterial host. The SV40 viral *ori* is a 65 bp region derived from the SV40 viral chromosome that when included in a nucleic acid sequence will allow replication of the nucleic acid in an animal cell. Inclusion of the SV40 *ori* region in a plasmid that also has the *E. coli ori* element will allow the plasmid to be replicated in both a bacterial host and in an animal cell.

The term "centromere" as used herein refers to a small, specialized region of a chromosome recognized as a constriction in a condensed chromosome. A kinetochore lies within the centromeric region and is attached to microtubules extending to the poles of a dividing cell.

The term "telomere" as used herein refers to repetitive oligomeric nucleic acid sequences located at the ends of linear eukaryotic chromosomes. Telomeres are required to prevent shortening of chromosomal DNA during replication of the linear nucleic acid.

Recombinant expression vectors can be designed for the expression of the encoded proteins eukaryotic cells. Useful vectors may comprise constitutive or inducible promoters to direct expression of either fusion or non-fusion proteins. With fusion vectors, a number of amino acids are usually added to the expressed target gene sequence such as, but not limited to, a protein sequence for thioredoxin. A proteolytic cleavage site may further be introduced at a site between the target recombinant protein and the fusion sequence. Additionally, a region of amino acids, such as a polymeric histidine region, may be introduced to allow binding of the fusion protein to metallic ions such as nickel bonded to a solid support, and thereby allow purification of the fusion protein. Once the fusion protein has been purified, the cleavage site allows the target recombinant protein to be separated from the fusion sequence. Enzymes suitable for use in cleaving the proteolytic cleavage site include, but are not limited to, Factor Xa and thrombin. Fusion expression vectors that may be useful in the present invention include pGex (Amrad Corp., Melbourne, Australia), pRIT5 (Pharmacia, Piscataway, NJ) and pMAL (New England Biolabs, Beverly, MA), that fuse glutathione S-transferase, protein A, or maltose E binding protein, respectively, to the target recombinant protein.

Expression of a foreign gene can be obtained using eukaryotic host cells such as, but not limited to, mammalian or avian cells. The use of eukaryotic host cells permit partial or complete post-translational modification such as, but not only, glycosylation and/or the formation of the relevant inter- or intra-chain disulfide bonds. Examples of vectors useful

for expression in the chicken *Gallus gallus* include pYepSecl as in Baldari *et al.*, E.M.B.O.J. 6, 229-234 (1987) and pYES2 (Invitrogen Corp., San Diego, CA), incorporated herein by reference in their entireties. Once the isolated DNA molecule of the present invention has been cloned into an expression system, it is ready to be incorporated into a host cell.

The terms "transformation" and "transfection" as used herein refer to the process of inserting a nucleic acid into a cell. Many techniques are well known to those skilled in the art to facilitate transformation or transfection of a nucleic acid into a prokaryotic or eukaryotic organism. These methods involve a variety of techniques, such as treating the cells with high concentrations of salt such as, but not only, a calcium or magnesium salt, an electric field, detergent, or liposome mediated transfection, to render the host cell competent for the uptake of the nucleic acid molecules, and by such methods as sperm-mediated and restriction-mediated integration.

The term "transfecting agent" as used herein refers to a composition of matter added to the genetic material for enhancing the uptake of heterologous DNA segment(s) into a eukaryotic cell, preferably an avian cell, and more preferably a chicken male germ cell. The enhancement is measured relative to the uptake in the absence of the transfecting agent. Examples of transfecting agents include adenovirus-transferrin-polylysine-DNA complexes. These complexes generally augment the uptake of DNA into the cell and reduce its breakdown during its passage through the cytoplasm to the nucleus of the cell. These complexes can be targeted to the male germ cells using specific ligands that are recognized by receptors on the cell surface of the germ cell, such as the c-kit ligand or modifications thereof.

Other preferred transfecting agents include, but are not limited to, lipofectin, lipfectamine, DIMRIE C, Supeffect, and Effectin (Qiagen), unifectin, maxifectin, DOTMA, DOGS (Transfectam; dioctadecylamidoglycylspermine), DOPE (1,2-dioleoyl-sn-glycero-3-phosphoethanolamine), DOTAP (1,2-dioleoyl-3-trimethylammonium propane), DDAB (dimethyl dioctadecytammonium bromide), DHDEAB (N,N-di-n-hexadecyl-N,N-dihydroxyethyl ammonium bromide), HDEAB (N-n-hexadecyIN,N-dihydroxyethylammonium bromide), polybrene, or poly(ethylenimine) (PEI). These nonviral agents have the advantage that they can facilitate stable integration of xenogeneic DNA sequences into the vertebrate genome, without size restrictions commonly associated with virus-derived transfecting agents.

The terms "intracytoplasmic sperm injection" and "ICSI" as used herein refer to delivering an exogenous nucleic acid to a recipient cell by associating the exogenous

nucleic acid with the head of a sperm cell and then delivering the sperm cell head to the recipient cell by microinjection. The exogenous nucleic acid may be integrated into the endogenous genomic nucleic acid of the sperm, non-integrated as an episomal element of the nucleic acid complement of the sperm head, or linked internally or externally to the head of the sperm. The terms "chICSI" and "CHICSI™" as used herein refer to intracytoplasmic sperm injection into a chicken cell.

The terms "sub-zonal injection" and "SUZI" refer to delivering viable spermatozoa to an oocyte by microinjection, wherein the sperm are microinjected between the zona pellucida and the cytoplasmic membrane of an oocyte.

The term "gene delivery (or transfection) mixture" as used herein, in the context of the methods of sperm mediated transfer described herein, refers to selected genetic material in an appropriate vector mixed, for example, with an effective amount of lipid transfecting agent, for example, a cationic or polycationic lipid, such as polybrene. The amount of each component of the mixture is chosen so that the genetic modification, *e.g.*, by transfection or transduction, of a specific species of male germ cell is optimized. Such optimization requires no more than routine experimentation. The ratio of DNA to lipid is broad, preferably about 1:1, although other proportions can also be utilized depending on the type of lipid transfecting agent used.

This application uses gene nomenclature accepted by the Cucurbit Genetics Cooperative as it appears in the *Cucurbit Genetics Cooperative Report* 18:85 (1995); herein incorporated by reference in its entirety. Using this gene nomenclature, genes are symbolized by italicized Roman letters. If a mutant gene is recessive to the normal type, then the symbol and name of the mutant gene appear in italicized lower case letters.

3.2 ABBREVIATIONS

Abbreviations used in the present specification include the following: aa, amino acid(s); bp, base pair(s); cDNA, DNA complementary to RNA; mRNA, messenger RNA; tRNA, transfer RNA; nt, nucleotide(s); SSC, sodium chloride-sodium citrate; MAR, matrix attachment region; DMSO, dimethyl sulfoxide; TPLSM, two photon laser scanning microscopy; REMI, restriction enzyme mediated integration; WEFs, whole embryo fibroblasts.

4. BRIEF DESCRIPTION OF THE FIGURES

FIGS. 1A-E illustrate the nucleotide sequence (SEQ ID NO: 6) comprising the chicken lysozyme gene expression control region (SEQ ID NO: 7), the nucleotide sequence

encoding the chicken expression optimized human interferon α2b (IFNMAGMAX; SEQ ID NO: 5) and a SV40 polyadenylation signal sequence (SEQ ID NO: 8).

FIG. 2 illustrates the nucleotide sequence SEQ ID NO: 5 encoding the chicken expression optimized human interferon α2b (IFNMAGMAX).

FIGS. 3A-E illustrate the nucleotide sequence SEQ ID NO: 7 encoding the chicken lysozyme gene expression control region.

FIG. 4 illustrates the nucleotide sequence SEQ ID NO: 8 encoding the SV40 polyadenylation signal sequence.

FIGS. 5A-C illustrate the nucleotide sequence SEQ ID NO: 9 encoding the chicken lysozyme 3' domain.

FIGS. 6A-J illustrate the nucleotide sequence SEQ ID NO: 10 encoding the lysozyme gene expression control region (SEQ ID NO: 7) linked to the insert having the nucleotide sequence of SEQ ID NO: 5 encoding the chicken expression-optimized human interferon α2b (IFNMAGMAX) and the chicken lysozyme 3' domain SEQ ID NO: 9.

FIG. 7 illustrates the nucleotide sequence SEQ ID NO: 11 of the combinatorial promoter MDOT.

FIGS. 8A-B illustrate the oligonucleotides and primers (SEQ ID NOS: 14-31) used in the formation of the chicken codon optimized human interferon α2b-encoding nucleic acid.

FIG. 9 illustrates the primers (SEQ ID NOS: 32-35) used in the synthesis of the MDOT promoter.

FIG. 10 illustrates the level of human monoclonal antibodies IgG expressed in the serum of transgenic chick using ELISA.

FIG. 11 illustrates the detection of EGFP positive bands from transgenic sperm.

5. DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to methods of introducing nucleic acids into avian oocytes by sperm-mediated transfection to produce a transgenic chicken or quail, or other avian species, carrying the transgene in the genetic material in all or most of its tissue, including germ-line tissue. The methods and vectors of the present invention further generate transgenic avians that express heterologous genes in the serum of the avian and/or are deposited into an avian egg, preferably in the egg white. Vectors containing promoters that direct high level of expression of the heterologous protein in the avian, particularly in the magnum for deposition into the avian egg are provided. Additional regulatory elements, such as MARs, IRES's, enhancers, polyadenlyation signals, etc., may be included in the vectors of the invention to improve expression and efficiency.

5.1 METHODS OF TRANSGENESIS

5.1.1 SPERM-MEDIATED INTEGRATION OF HETEROLOGOUS TRANSGENES

The transgenic avians of the present invention are most preferably generated using sperm-mediated transfection of nucleic acid into avian oocytes. Specifically, the present invention provides methods for introducing nucleic acids containing a transgene, preferably, a nucleic acid vector of the invention as described in Section 5.2, *infra*, into an avian oocyte by sperm-mediated transfection. In preferred embodiments, the nucleic acid is first introduced into an avian sperm *in vitro* by lipofection, electroporation, restriction enzyme mediated integration (REMI) or similar methods, or *in vivo* by microinjection into the testis, and the modified avian sperm is then delivered to an avian oocyte by natural coitus after the modified avian sperm are returned to the testis of a male avian or in the method in which the nucleic acid has been injected directly into the testis or *in vitro* by microinjection, intracytoplasmicosperm injection (ICSI) or artificial insemination of oocytes isolated from an ovulating female bird, thereby generating a transgenic zygote and chick. In certain embodiments, the male germ cells are irradiated, more preferably irradiated by gamma rays, before the heterologous nucleic acid is incorporated therein. In other embodiments, the testis is depopulated of sperm prior to introduction of the transfected sperm.

The present invention contemplates that any technique capable of transferring heterologous material into sperm could be used so long as the technique preserves enough of the sperm's fertilization functions, such that the resultant sperm will be able to fertilize the oocyte. It is understood that the heterologous nucleic acid may be integrated into the genome of a recipient cell such as a spermatogonial cell or a spermatogonial precursor cell

for subsequent transfer to an embryo or the testicular material of the recipient male animal, preferably a chicken. It is further understood that the heterologous nucleic acid may not be integrated into the genome of the recipient cell but delivered as an episome which may or may not be integrated into the genome of the resulting zygote or chick.

5

5.1.1.1 PREPARATION OF TRANSGENIC CONSTRUCT

One aspect of the present invention relates to the preparation of a transgene which is to be incorporated into the genome of an avian sperm. In certain embodiments, the transgene comprises at least one heterologous nucleic acid. It is contemplated to be within 10 the scope of the present invention for the heterologous nucleic acid to comprise an expression vector such as, but not limited to, viral vectors, plasmid vectors, or linearized nucleic acid vectors or a combination thereof. (See section 5.2, *infra*, for details on vectors, and the preparation thereof). The expression vector may particularly be any suitable nonviral vector including plasmid DNA, bacteria artificial chromosomes (BACs), yeast 15 artificial chromosomes (YACs), etc. The expression vector may also be any suitable viral vector, for example, retroviral vectors, adenoviral vectors, transferrin-polylysine enhanced adenoviral vectors, human immunodeficiency virus vectors, lentiviral vectors, Moloney murine leukemia virus-derived vectors, and virus-derived DNAs that facilitate polynucleotide uptake by and release into the cytoplasm of germs cells.

20 Transcriptional promoters of an expression vector of the present invention may be a constitutively active promoter such as the cytomegaloviral promoter or Rous sarcoma virus promoter, or a tissue-specific promoter, preferably a tissue-specific promoter operable in oviduct cells of an avian species including, but not limited to, the promoters of the genes encoding ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin. 25 Optionally, the transcriptional promoter of an expression vector may be a regulatable promoter. The expression vector may further comprise a region encoding a transcriptional terminator, such as a bovine growth hormone transcriptional terminator.

In preferred embodiments, a transgene construct comprises at least two separate or independent elements. A first element could comprise an oviduct-specific promoter, such 30 as, but not limited to ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin, which would drive expression of a gene coding for a protein of interest in the oviduct. A second element can be located either upstream or downstream for the first element and comprises a protamine promoter, or a segment thereof that is sufficient to drive the expression of a marker gene such as the Green Fluorescent Protein (GFP) to facilitate 35 identification of transfected sperm.

In one embodiment of the present invention, the heterologous nucleic acid comprises cohesive ends characterized as capable of hybridizing to cohesive ends generated by a restriction endonuclease. The cohesive ends on the nucleic acid may be generated by restriction endonuclease cleavage of a circular or linear nucleic acid, by the chemical addition of nucleotides to the ends of a linear nucleic acid, or by a combination of chemical and enzymatic methods.

In another embodiment of the present invention, the heterologous nucleic acid is linearized and has at least one blunt end. The blunt end of the nucleic acid may be generated, by an exonuclease digestion of cohesive ends, such as S1 nuclease.

In the methods of generating transgenic cells according to the present invention, the genomic nucleic acid of the recipient cell, male germ cell or oocyte can be cleaved to receive the integrating heterologous nucleic acid. Any method may be selected that will generate limited, random cleavage that will allow integration of the heterologous nucleic acid into the genome of the recipient cell or oocyte. When the integrating heterologous nucleic acid has cohesive ends, the recipient genomic nucleic acid may be cleaved with a restriction endonuclease generating cohesive ends capable of hybridizing to the cohesive ends of the heterologous nucleic end. When the heterologous nucleic acid has blunt ends, the genomic nucleic acid can be cleaved by any method that will generate blunt ends at the cleavage site, including restriction endonuclease cleavage, or irradiation of the cell with high-energy irradiation. Suitable radiations that may be applied to the methods of the present invention include, for example, gamma rays, x-rays, ultraviolet light or ultrasound. It is contemplated that the cleavage of genomic nucleic acid and integration of a heterologous nucleic acid therein will result in a viable recipient cell that can be used to fertilize an avian oocyte, or will not yield a viable cell. A non-viable sperm cell may, however, be used to deliver the transgene to an oocyte using, for example, the ICSI (CHICSI™) method.

The heterologous nucleic acid of the present invention may further comprise a centromere element and at least one telomere element. In one embodiment, the centromere and the at least one telomeres are derived from the chicken. While the *ori* site alone will allow replication of the heterologous nucleic acid when transfected into an oocyte or zygote thereof, segregation of the replicates into each daughter cell will require the optional centromeric element. In the absence of this centromeric element, segregation will be random between daughter cells with some daughter cells not receiving one copy of the transgenic nucleic acid. A mosaic transgenic animal would, therefore, result.

In one embodiment of the present invention, therefore, the heterologous nucleic acid is an artificial chromosome comprising a heterologous transgenic element having the properties desired to be expressed by a transgenic animal, an origin of replication (*ori*) site, and a centromere. In this embodiment, the heterologous nucleic acid may be a circular nucleic acid or a linear nucleic acid. In another embodiment, the heterologous nucleic acid is a linear nucleic acid further comprising telomeres.

In another aspect of the methods according to the present invention, the transgenic oocyte or ovum of the present invention is incubated for development of the zygote therein to a fetus, and subsequently to a chick for hatching. In one embodiment of the present invention, therefore, the zygote is incubated in a surrogate avian female, wherein the method comprises the steps of fistulating an avian female, delivering the avian oocyte to the infundibulum of the female bird, allowing the avian female to incubate the avian oocyte to an embryo within an egg, allowing the avian female to lay the egg, and allowing the embryo to hatch as a viable chick, wherein the chick is a transgenic chick having an exogenous nucleic acid incorporated therein.

5.1.1.2 SPERM TRANSGENESIS

The heterologous nucleic acid may be delivered to an avian male germ cell (*i.e.*, sperm, spermatozoon cell or a precursor cell) by a method such as by contacting the male germ cell with a gene delivery mixture comprising a nucleic acid, either a eukaryotic viral vector or a vector that is not derived from a eukaryotic virus, at about or below the avian's body temperature and for an effective period of time such that the nucleic acid is incorporated into the cell, and preferably into the genome of the cell, optionally isolating or selecting the genetically modified cell with the aid of a genetic selection marker expressed in the genetically modified cell, transferring the isolated or selected genetically modified germ cell to a testis of a recipient male avian such that the cell lodges in a seminiferous tubule of the testis. A genetically modified male gamete may be produced therein, and breeding the recipient male avian with a female avian of its species will generate transgenic progeny that carry the heterologous transgenic nucleic acid in its genome.

In certain embodiments, the avian male germ cells are isolated and removed from a male avian. The avian male germ cells is then transfected by introducing the heterologous nucleic acid into the genome of the avian male germ cells by lipofection, electroporation, restriction enzyme mediated infection (REMI) or similar methods. In certain other embodiments, the heterologous nucleic acid is injected directly into the testis of the male avian for transfection. Male germ cells can be extracted to determine whether transfection

has occurred or the extent of transfection. The male avian can be mated with a female avian to produce transgenic offsprings or the sperm can be used for IVF.

The precursor cell may be selected from the group consisting of spermatogonial stem cells, type B spermatogonia, primary spermatocytes, preleptotene spermatocytes, leptotene spermatocytes, zygotene spermatocytes, pachytene spermatocytes, secondary spermatocytes, and spermatids. The embodiment further comprises the steps of incorporating the heterologous transgene into the genome of the spermatozoon cell or the precursor cell, so that a genetically modified male gamete is produced by the male avian, and breeding the male avian with a female of the same species such that a transgenic progeny is thereby produced that carries the polynucleotide in its genome.

In certain embodiments, the heterologous genetic material may be introduced into the genome of an avian male germ cell, such that a polynucleotide is delivered using known gene delivery systems to male germ cells *in situ* in the testis of the male avian (*e.g.*, by *in vivo* transfection or transduction). In one embodiment, the invention relates to an *in vitro* method of incorporating heterologous genetic material into the genome of a male avian by isolating male germ cells *ex corpora*, delivering a polynucleotide thereto, and then returning the transfected cells to the testes of a recipient male bird. In yet another embodiment, the *in vitro* method involves microinjecting the recombinant male germ cells into a recipient fertilized oocyte, whereupon the sperm head enters the oocyte nucleus to deliver the heterologous nucleic acid thereto.

In a preferred embodiment, the invention relates to an *in vivo* method that injects a gene delivery mixture, preferably into the seminiferous tubules, or into the testis, and most preferably into the vas efferens or vasa efferentia using, for example, a micropipette and a picopump delivering a precise measured volume under controlled amounts of pressure. The modified germ cells differentiate in their own milieu. Progeny animals exhibiting the nucleic acid's integration into its germ cells (*i.e.*, transgenic animals) are selected. The selected progeny can then be mated, or their sperm utilized for insemination or *in vitro* fertilization, to produce further generations of transgenic progeny or for microinjection into isolated oocytes.

In another preferred embodiment, the invention relates to an *in vitro* method wherein male germ cells are obtained or collected from a donor male avian, by any means known in the art such as, for example, transection of the testes. The male germ cells are then exposed to a gene delivery mixture, preferably within several hours of collection, or cryopreserved for later use. When the male germ cells are obtained from the donor avian by transection of the testes, the cells can be incubated in an enzyme mixture known for gently breaking up the

tissue matrix and releasing undamaged cells. Suitable enzymes to disrupt the integrity of a tissue include, but are not limited to, pancreatic trypsin, collagenase type I, pancreatic DNAse type I, as well as bovine serum albumin and a modified DMEM medium. After washing the cells, they can be placed in an incubation medium such as DMEM or the like, and plated on a culture dish for genetic modification by exposure to a gene delivery mixture.

In other embodiments, a transgene can be incorporated into an avian sperm by lipofection, electroporation, restriction enzyme mediated integration (REMI), intracytoplasmic sperm injection (ICSI) or similar methods.

Liposome

In a preferred embodiment, a transgene is incorporated into an avian sperm by liposomes. The male germ cells, which may be intact and viable spermatozoa, or the non-viable heads thereof, may be transfer to a recipient oocyte using liposome-mediated delivery. PCT Publication WO 87/05325, which is incorporated by reference herein in its entirety, discloses a method of transferring organic and/or inorganic material into sperm or egg cells by using liposomes. The heterologous nucleic acid can also be incorporated into a male sperm using Lipofectin-based liposomes. (*See, e.g.*, Bachiller *et al.*, 1991, *Mol. Reprod. Develop.* 30: 194-200; Nakanishi and Iritani, 1993, *Mol. Reprod. Develop.* 36: 258-261).

Electroporation

In another preferred embodiment, a transgene is incorporated into an avian sperm by electroporation. The application of electrical current has been shown to enhance the uptake of exogenous DNA fragments by cultured cells. Enhancement of nuclear uptake of the heterologous DNA will promote earlier chromosomal integration of the exogenous DNA molecules, thus reducing the degree of genetic mosaicism observed in transgenic avian founders.

In one embodiment, the male germ cells is placed in a cuvette and a solution of the transgenic nucleic acid coding the protein of interest is added. A direct current pulse is discharged in the cuvette suspension. The current pulse creates temporary, short-lived pores in the cell membrane and allow the male germ cells to take up the transgene while only slightly compromising cell viability. More description on the use of electroporation to incorporate DNA can be found in Gagne *et al.*, 1991, *Mol. Reprod. Develop.* 29: 6-15, which is incorporated herein by reference in its entirety.

Restriction Enzyme Mediated Integration (REMI)

In yet another preferred embodiment, a transgene is incorporated into an avian sperm by restriction enzyme mediated integration (REMI). The heterologous nucleic acid to be integrated into, for example, the sperm nuclear DNA is converted to a linear double stranded DNA possessing single-stranded cohesive ends by contacting the heterologous DNA with a type II restriction enzyme that upon scission, generates such ends. The nucleic acid to be cut can be a circular nucleic acid such as in a plasmid or a viral vector or a linear nucleic acid that possesses at least one recognition and cutting site outside of the genes or regulatory regions critical to the desired post-integration function of the nucleic acid, and no recognition and cutting sites within the critical regions.

Alternatively the heterologous DNA to be integrated into the sperm nuclear DNA can be prepared by chemically and/or enzymatically adding cohesive ends to a linear DNA. The added cohesive ends must be able to hybridize to the cohesive ends characteristic of a nucleic acid cleaved by a type II restriction endonuclease. Alternatively, the cohesive ends can be added by combining the methods based on type II restriction enzyme cutting and chemical and/or enzymatic addition. It is also within the scope of the present invention for the linearized nucleic acid to have one end that is a blunt end without unpaired nucleotides. Such blunt ends can be generated by restriction endonuclease digestion, exonuclease digestion of cohesive ends or fill-in of cohesive ends by polynucleotide synthesis, using methods as described, for example, in Sambrook *et al.*, (*supra*), incorporated herein by reference in its entirety.

It is also to be understood that a nucleic acid to be delivered to a recipient cell may be cleaved with two different restriction endonucleases that may generate the same or different cohesive termini, or at least one blunt-end terminus. Neither restriction endonucleases will have a recognition site within the nucleic acid sequence required to be a transgene in the recipient cell.

When a restriction endonuclease is used to cleave the genomic nucleic acid of the recipient cell, the endonuclease may be co-delivered to the recipient cell such as a sperm cell with the heterologous nucleic acid, or sequentially delivered. If a nucleic acid is cleaved with at least two restriction endonucleases, thereby generating at least one cohesive terminus, the at least two endonucleases may be delivered to a recipient cell either together or sequentially. The transfected nucleic acid may be mixed with at least one of the endonucleases or delivered to a recipient cell before or after at least one endonuclease is delivered thereto.

At least one terminus of a linearized nucleic acid to be delivered to a recipient cell may be a blunt end terminus, generated by endonuclease cleavage, chemical synthesis, enzyme directed nucleic acid digestion or synthesis, or any combination thereof. A recipient cell genome such as a sperm cell genome, may therefore be cleaved before, during or after delivery of the linearized nucleic acid to the cell, by delivery of a blunt-end generating restriction endonuclease to the recipient cell, or by radiation-induced cleavage. Suitable radiations that may be applied to, for example, a sperm cell include, but are not limited to, gamma radiation, x-rays, ultraviolet light and ultrasound. The dose and duration of the radiation applied to a cell sample are determined for each sample, for levels of cleavage that will allow integration of the transfected nucleic acid into the cell genome, while maintaining viability of the cells for use in artificial insemination or recolonization of an avian testes. Viability of a recipient sperm may not be required when the transfected sperm are delivered to a recipient avian oocyte by such procedures as ICSI or CHICSI™. Cleavage of the genomic nucleic acid by irradiation or ultrasound can be either before, during or after delivery of the heterologous nucleic acid to the recipient cell.

While not wishing to be bound by any one theory, the transfected nucleic acid may be integrated into a cleavage site of the genomic nucleic acid. Integration may be facilitated by the cohesive ends on the heterologous nucleic acid that hybridize to the like cohesive ends of the cleaved genomic nucleic acid. The integrated heterologous nucleic acid will then replicate and segregate with the genome of the recipient cell.

Alternatively, the heterologous nucleic acid may not be integrated into a recipient genome, but will remain as an extrachromosomal episome. The heterologous nucleic acid of the present invention may circularize by hybridization of the cohesive ends of the nucleic acid, rather than be integrated into the genome. When the heterologous nucleic acid comprises any natural or synthetic origin of replication (*ori* element) the nucleic acid will be capable of replicating independently of the recipient genome. In one embodiment of the present invention the *ori* site included with a heterologous nucleic acid is derived from the SV40 virus. Episomal replication and segregation of daughter copies of the episome is facilitated by the linearized viral *ori* site and/or a centromere isolated from, for example, a chicken chromosome, thereby generating a chicken artificial chromosome. In another embodiment, the linearized heterologous nucleic acid will not be integrated into the genome of the recipient cell but remain as a separate unit that, because of a centromeric structure incorporated therein, will segregate into daughter cells during mitotic division. In this case, the unincorporated episomal heterologous nucleic acid is a chicken artificial chromosome (CAC).

The REMI method for stably integrating heterologous DNA into the genomic DNA of a recipient cell is described by Shemesh *et al.* in PCT Publication No. WO 99/42569 and incorporated herein by reference in its entirety. This REMI method comprises in part an adaptation of the REMI technique disclosed by Schiest and Petes (Proc. Nat. Acad. Sci. U.S.A. 88, 7585-7589 (1991)) and Kuspa and Loomis (Proc. Nat. Acad. Sci. U.S.A., 89, 8803-8807 (1992)) both incorporated herein by reference in their entireties.

In preferred embodiments, the avian sperm are irradiated before being exposed to gene delivery mixture or having a transgene incorporated therein. The male germ cells can be irradiated with a suitable dose of gamma irradiation, preferably, 1 Gy, 2 Gy, 3 Gy, 4 Gy, 5 Gy, 6 Gy, 7 Gy, 8 Gy, 9 Gy, 10 Gy, 11 Gy, 12 Gy, 15 Gy or 20 Gy, without compromising the viability and/or mobility of the sperms. (See Wooster *et al.*, 1977, *Can. J. Genet. Cytol.* 19: 437-446).

Whether employed in the *in vivo, in situ* or *in vitro* method, the gene delivery mixture, once in contact with the male germ cells, facilitates the uptake and transport of heterologous genetic material into the appropriate cell location for integration into the genome and expression. A number of known gene delivery methods can be used for the uptake of nucleic acid sequences into the cell and facilitate the integration of the heterologous nucleic acid into the genome of the recipient cell. Such methods include, but are not limited to viral vectors, liposomes, electroporation, REMI, and ICSI.

A gene delivery mixture suitable for use in the *in vivo, in situ* or *in vitro* methods of sperm-mediated transfection comprises a nucleic acid encoding a desired trait or product, and a suitable promoter sequence such as, for example, a tissue-specific promoter, or an IRES. The transgenic nucleic acids of the present invention may further comprise an origin of replication. For example, an origin of replication may be the SV40 *ori*, or a centromere derived from the chicken. A linear nucleic acid may further comprise a telomere at one or both ends of the nucleic acid.

Optionally, agents that increase the uptake of, or comprise non-eukaryotic viral vectors, *e.g.*, plasmids, BACs, YACs, etc., the nucleic acid sequence, such as liposomes, retroviral vectors, adenoviral vectors, adenovirus enhanced gene delivery systems, or combinations thereof may be included in the gene delivery mixture. A reporter construct, including a genetic selection marker, such as the gene encoding for Green Fluorescent Protein, may also be added to the gene delivery mixture. Targeting molecules, such as c-kit ligand, can be added to the gene delivery mixture to enhance the transfer of genetic material into the male germ cell. An immunosuppressing agent, such as cyclosporin or a corticosteroid may also be added to the gene delivery mixture as known in the art.

Any of a number of commercially available gene delivery mixtures can be used, to which the polynucleotide encoding a desire trait or product is further admixed. The final gene delivery mixture comprising the polynucleotide can then be admixed with the male gamete cells and allowed to interact for a period of between about 2 hours to about 16 hours, at a temperature of about 33°C to about 37°C. After this period, the cells are preferably placed at a lower temperature of about 33°C to about 34°C, for about 4 hours to about 20 hours, preferably about 16 to about 18 hrs.

Isolating and/or selecting genetically transgenic germ cells (and transgenic somatic cells, and of transgenic vertebrates) is by any suitable means, such as, but not limited to, physiological and/or morphological phenotypes of interest using any suitable means, such as biochemical, enzymatic, immunochemical, histologic, electrophysiologic, biometric or like methods, and analysis of cellular nucleic acids, for example the presence or absence of specific DNAs or RNAs of interest using conventional molecular biological techniques, including hybridization analysis, nucleic acid amplification including, but not limited to, polymerase chain reaction, transcription-mediated amplification, reverse transcriptase-mediated ligase chain reaction, and/or electrophoretic technologies.

One preferred method of isolating or selecting male germ cell populations comprises obtaining specific male germ cell populations, such as spermatogonia, from a mixed population of testicular cells by extrusion of the cells from the seminiferous tubules and enzyme digestion. The spermatogonia, or other male germ cell populations, can be isolated from a mixed cell population by methods such as the utilization of a promoter sequence that is specifically or selectively active in cycling male germ line stem cell populations. Suitable promoters include B-Myb or a specific promoter, such as the c-kit promoter region, c-raf-1 promoter, ATM (ataxia-telangiectasia) promoter, vasa promoter, RBM (ribosome binding motif) promoter, DAZ (deleted in azoospermia) promoter, XRCC-1 promoter, HSP 90 (heat shock gene) promoter, cyclin A1 promoter, or FRMI (from Fragile X site) promoter and the like. A selected promoter may be linked to a reporter construct, for example, a construct comprising a gene encoding Green Fluorescent Protein (or EGFP), Yellow Fluorescent Protein, Blue Fluorescent Protein, a phycobiliprotein, such as phycoerythrin or phycocyanin, or any other protein which fluoresces under suitable wave-lengths of light, or encoding a light-emitting protein, such as luciferase or apoaequorin. The unique promoter sequences drive the expression of the reporter construct only during specific stages of male germ cell development (*e.g.*, Mailer *et al.*, 1999, *J. Biol. Chem.* 276(16), 11220-28; Schrans-Stassen *et al.*, 1999, *Endocrinology* 140, 5894-5900, both of which are incorporated herein by reference in their entireties). In the case of a fluorescent reporter construct, the cells can be

sorted with the aid of, for example, a FACS set at the appropriate wavelength(s), or they can be selected by chemical methods.

Male germ cells that have the DNA modified in the desired manner are isolated or selected, and transferred to the testis of a suitable recipient avian, preferably the donor male avian of the male germ cells. Further selection can be attempted after biopsy of one or both of the recipient male's testes, or after examination of the animal's ejaculate amplified by the polymerase chain reaction to confirm that the desired nucleic acid sequence had been incorporated.

The genetically modified germ cells isolated or selected as described above are transferred to a testis of a suitable male avian, preferably a chicken, that can be, but need not be, the same donor animal. Before transferring the genetically modified male germ cells to the recipient animal, the testes of the recipient are depopulated of endogenous germ cells, thereby facilitating the colonization of the recipient testis by the genetically modified germ cells. Depopulation of the testis has commonly been accomplished by exposing the whole animal to gamma irradiation or by localized irradiation of the testis. The basic rigid architecture of the gonad should not be destroyed, nor significantly damaged. Disruption of tubules may lead to impaired transport of testicular sperm and result in infertility. Sertoli cells should not be irreversibly damaged, as they provide a base for development of the germ cells during maturation, and for preventing the host immune defense system from destroying grafted foreign spermatogonia.

Suitable denuding methods, include irradiation by gamma-rays, x-rays, ultrasound, ultraviolet light, by chemical treatment, by means of infectious agents such as viruses, or by autoimmune depletion or by combinations thereof, preferably by a combined treatment of the vertebrate with an alkylating agent and gamma irradiation as taught in WO 00/69257, incorporated herein by reference in its entirety.

Gamma radiation-induced spermatogonial degeneration probably related to the process of apoptosis. (Hasegawa *et al.*, 1998, *Radiat. Res.* 149: 263-70). Alternatively, a composition containing an alkylating agent such as busulfan (MYLERAN™) can be used to depopulate. (Jiang F.X., 1998, *Anat. Embryol.* 198(1): 53-61; Russell and Brinster, 1996, *J. Androl.* 17(6): 615-27; Boujrad *et al.*, 1995, *Andrologia* 27(4): 223-28; Linder *et al.*, 1992, *Reprod. Toxicol.* 6(6): 491-505; Kasuga and Takahashi, 1986, *Endocrinol. Jpn* 33(1): 105-15). Other cytotoxic alkylating agent, may be, but is not limited to, chlorambucil, cyclophosphamide, melphalan, or ethyl ethanesulfonic acid, and may be combined with gamma irradiation, to be administered in either sequence.

The dose of the alkylating agent and the dose of gamma radiation are in an amount sufficient to substantially depopulate the testis. The alkylating agent can be administered by any pharmaceutically acceptable delivery system, including but not limited to, intraperitoneal, intravenous, or intramuscular injection, intravenous drip, implant, transdermal or transmucosal delivery systems.

The isolated or selected genetically modified germ cells are transferred into the recipient testis by direct injection using a suitable micropipette. Support cells, such as Leydig or Sertoli cells, that can be unmodified or genetically modified, can be transferred to a recipient testis along with the modified germ cells.

5.1.1.3 DELIVERY OF TRANSGENIC SPERM TO OOCYTES

The transfected male avian germ cells may be used to deliver a heterologous nucleic acid to an avian oocyte by implanting the transfected male germ cells such as transfected spermatogonial precursor cells, into the testicular tissue of host male birds previously denuded of viable spermatogonial cells or sperm. The implanted transfected male avian germ cells may colonize the testicular tissue, proliferate therein, and generate viable transgenic sperm that may be harvested for use in artificial insemination procedures, or transferred to a recipient oocyte by natural coitus.

In certain embodiments, therefore, the transgenic avian may be produced by the sperm-mediated transfer of at least one heterologous transgene. The transgene may be incorporated into the genomic nucleic acid of a spermatozoon cell or a precursor thereof, so that a genetically modified avian sperm is produced by the male avian. Breeding the male avian with a female of its species will generate a transgenic progeny carrying the at least one transgene in its genome.

A union of male and female gametes to form a transgenic zygote is brought about by copulation of the male and female vertebrates of the same species, or by *in vitro* or *in vivo* artificial means. If artificial means are chosen, then incorporating into the genome a genetic selection marker that is expressed in male germ cells is particularly useful.

Suitable artificial means include, but are not limited to, artificial insemination, *in vitro* fertilization (IVF) and/or other artificial reproductive technologies, such as intracytoplasmic sperm injection (ICSI), subzonal insemination (SUZI), or partial zona dissection (PZD). Also others, such as cloning and embryo transfer, cloning and embryo splitting, and the like, can be employed.

In a preferred embodiment, a transgene is incorporated into an avian sperm by intracytoplasmic sperm injection (ICSI). The male germ cells, which may be intact and

viable spermatozoa, or the non-viable heads thereof, may be microinjected into the cytoplasm or the nucleus of an isolated oocyte such as an avian oocyte, preferably a chicken oocyte, by any method known to one of skill in the art, including, for example, combining a confocal microscope and micromanipulator, or the like to visualize and monitor the microinjection of an opaque avian egg.

The transgenic vertebrate progeny can, in turn, be bred by natural mating, artificial insemination, or by *in vitro* fertilization (IVF) and/or other artificial reproductive technologies, such as intracytoplasmic sperm injection (ICSI) and chicken intracytoplasmic sperm injection (CHICSI™), subzonal insemination (SUZI), or partial zona dissection (PZD), to obtain further generations of transgenic progeny. Although the genetic material is originally inserted solely into the germ cells of a parent animal, it will ultimately be present in the germ cells of future progeny and subsequent generations thereof. In addition, the genetic material will also be present in cells of the progeny other than germ cells, *i.e.*, somatic cells.

The methods of the present invention may further comprise returning a transfected fertilized oocyte to a surrogate mother, especially a female chicken, for the continued incubation and development of the transgenic zygote. With chickens, the developed embryo is laid as a hard-shell egg that will hatch as a viable chick. When the heterologous nucleic acid is directly integrated into the genome of the oocyte, the transgenic chick will include the transgenic heterologous nucleic acid in all of its cells. Where the heterologous nucleic acid is episomal with respect to the genome of the transgenic zygote and chick, and the episomal nucleic acid comprises a centromeric body, most, if not all, of the cells of the zygote and chick will comprise the heterologous nucleic acid. When the episomal nucleic acid does not include a centromeric body, however, the transgenic zygote and chick can be a mosaic wherein expression of the exogenous transgene will only occur in some, but not all cells or tissues of the transgenic animal.

5.1.2 BREEDING AND MAINTENANCE OF TRANSGENIC AVIAN

Another aspect of the present invention is a transgenic avian produced by the methods of the present invention and producing a heterologous polypeptide in an egg, wherein the transgenic avian comprises at least one heterologous nucleic acid sequence encoding the polypeptide and wherein the heterologous polypeptide is delivered to the white of an avian egg by a female of the avian.

The invention relates to a method of producing transgenic avians that express significant quantities of useful heterologous proteins, *e.g.*, therapeutic and diagnostic

proteins, including immunoglobulins, industrially useful proteins and other biologics etc. in the avian egg white. The heterologous protein can then be readily purified from the avian egg. The methods of the invention provide improved efficiencies of transgenesis, transmission of the transgene and/or level of heterologous protein expression. Another aspect of the invention is a method of producing a transgenic avian capable of expressing a heterologous protein. Therefore, the present invention relates to methods of producing transgenic avians, preferably chickens, wherein the incorporated transgene may be expressed as a constituent protein of the white of a hard-shell egg.

Although the genetic material is originally inserted solely into the germ cells of a parent animal, it will ultimately be present in the germ cells of future progeny and subsequent generations thereof. In addition, the genetic material will also be present in cells of the progeny other than germ cells, *i.e.*, somatic cells.

Using the methods of the invention for producing transgenic avians, particularly methods using vectors that are not derived from eukaryotic viruses, and, preferably, the methods of cytoplasmic micro-injection described herein, the level of mosaicism of the transgene (percentage of cells containing the transgene) in avians hatched from microinjected embryos (*i.e.*, the G_0s) is greater than 5%, 10%, 25%, 50%, 75% or 90%, or is the equivalent of one copy per one genome, two genomes, five genomes, seven genomes or eight genomes, as determined by any number of techniques known in the art and described *infra*.

In additional particular embodiments, the percentage of G0s that transmit the transgene to progeny (G1s) is greater than 5%, preferably, greater than 10%, 20%, 30%, 40%, and, most preferably, greater than 50%, 60%, 70%, 80%, 90%. In other embodiments, the transgene is detected in 10%, 20%, 30%, 40%, and most preferably, greater than 50%, 60%, 70%, 80%, 90% of chicks hatching from embryos into which nucleic acids have been introduced using methods of the invention.

5.2 VECTORS

A variety of vectors useful in carrying out the methods of the present invention are described herein. These vectors may be used for stable introduction of a selected heterologous polypeptide-coding sequence (and/or regulatory sequences) into the genome of an avian, in particular, to generate transgenic avians that produce exogenous proteins in specific tissues of an avian, and in the oviduct in particular, or in the serum of an avian. In still further embodiments, the vectors are used in methods to produce avian eggs containing exogenous protein.

In particular embodiments, preferably for use in the sperm-mediated transgenesis methods described herein, the vectors of the invention are not derived from eukaryotic viral vectors or retroviral vectors (except in certain embodiments for containing eukaryotic viral regulatory elements such as promoters, origins of replication, etc). In particular embodiments, the vector is not an REV, ALV or MuLV vector. In particular, useful vectors include, bacteriophages such as lambda derivatives, such as λgt11, λgt WES.tB, Charon 4, and plasmid vectors such as pBR322, pBR325, pACYC177, pACYC184, pUC8, pUC9, pUC18, pUC19, pLG339, pR290, pKC37, pKC101, SV40, pBLUESCRIPT® II SK +/- or KS +/- (see "Stratagene Cloning Systems" Catalog (1993) from STRATAGENE®, La Jolla, Calif., which is hereby incorporated by reference), pQE, pIH821, pGEX, pET series (see Studier, F.W. *et. al.*, 1990, "Use of T7 RNA Polymerase to Direct Expression of Cloned Genes" *Gene Expression Technology* 185, which is hereby incorporated by reference) and any derivatives thereof, cosmid vectors and, in preferred embodiments, artificial chromosomes, such as, but not limited to, YACs, BACs, BBPACs or PACs. Such artificial chromosomes are useful in that a large nucleic acid insert can be propagated and introduced into the avian cell.

In other particular embodiments, as detailed above in section 5.2, *infra*, the vectors of the invention are derived from eukaryotic viruses, preferably avian viruses, and can be replication competent or, preferably, replication deficient. In particular embodiments, the vectors are derived from REV, ALV or MuLV. Nucleic acid sequences or derivative or truncated variants thereof, may be introduced into viruses such as vaccinia virus. Methods for making a viral recombinant vector useful for expressing a protein under the control of the lysozyme promoter are analogous to the methods disclosed in U.S. Patent Nos. 4,603,112; 4,769,330; 5,174,993; 5,505,941; 5,338,683; 5,494,807; 4,722,848; Paoletti, E., 1996, *Proc. Natl. Acad. Sci.* 93: 11349-11353; Moss, 1996, *Proc. Natl. Acad. Sci.* 93: 11341-11348; Roizman, 1996, *Proc. Natl. Acad. Sci.* 93: 11307-11302; Frolov *et al.*, 1996, *Proc. Natl. Acad. Sci.* 93: 11371-11377; Grunhaus *et al.*, 1993, *Seminars in Virology* 3: 237-252 and U.S. Patent Nos. 5,591,639; 5,589,466; and 5,580,859 relating to DNA expression vectors, *inter alia*; the contents of which are incorporated herein by reference in their entireties.

Recombinant viruses can also be generated by transfection of plasmids into cells infected with virus.

Preferably, vectors can replicate (*i.e.*, have a bacterial origin of replication) and be manipulated in bacteria (or yeast) and can then be introduced into avian cells. Preferably, the vector comprises a marker that is selectable and/or detectable in bacteria or yeast cells

and, preferably, also in avian cells, such markers include, but are not limited to, Amp^r, tet^r, LacZ, etc. Preferably, such vectors can accommodate (*i.e.*, can be used to introduce into cells and replicate) large pieces of DNA such as genomic sequences, for example, large pieces of DNA consisting of at least 25 kb, 50 kb, 75 kb, 100 kb, 150 kb, 200 kb or 250 kb, such as BACs, YACs, cosmids, etc.

The insertion of a DNA fragment into a vector can, for example, be accomplished by ligating the DNA fragment into a vector that has complementary cohesive termini. However, if the complementary restriction sites used to fragment the DNA are not present in the vector, the ends of the DNA molecules may be enzymatically modified. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated linkers may comprise specific chemically synthesized oligonucleotides encoding restriction endonuclease recognition sequences. In an alternative method, the cleaved vector and the transgene may be modified by homopolymeric tailing.

The vector can be cloned using methods known in the art, *e.g.*, by the methods disclosed in Sambrook *et al.*, (*supra*); Ausubel *et al.*, 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y., both of which are hereby incorporated by reference in their entireties. Preferably, the vectors contain cloning sites, for example, restriction enzyme sites that are unique in the sequence of the vector and insertion of a sequence at that site would not disrupt an essential vector function, such as replication.

As discussed above, vectors used in certain methods of the invention preferably can accommodate, and in certain embodiments comprise, large pieces of heterologous DNA such as genomic sequences, particularly avian genomic sequences. Such vectors can contain an entire genomic locus, or at least sufficient sequence to confer endogenous regulatory expression pattern, *e.g.*, high level of expression in the magnum characteristic of ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin, etc, and to insulate the expression of the transgene sequences from the effect of regulatory sequences surrounding the site of integration of the transgene in the genome. Accordingly, as detailed below, in preferred embodiments, the transgene is inserted in an entire genomic loci or significant portion thereof.

To manipulate large genomic sequences contained in, for example, a BAC, nucleotide sequences coding for the heterologous protein to be expressed and/or other regulatory elements may be inserted into the BAC by directed homologous recombination in bacteria, *e.g.*, the methods of Heintz WO 98/59060; Heintz *et al.*, WO 01/05962; Yang *et*

al., 1997, *Nature Biotechnol.* 15: 859-865; Yang *et al.*, 1999, *Nature Genetics* 22: 327-35; which are incorporated herein by reference in their entireties.

Alternatively, the BAC can also be engineered or modified by "E-T cloning," as described by Muyrers *et al.* (1999, *Nucleic Acids Res.* 27(6): 1555-57, incorporated herein by reference in its entirety). Using these methods, specific DNA may be engineered into a BAC independently of the presence of suitable restriction sites. This method is based on homologous recombination mediated by the recE and recT proteins ("ET-cloning") (Zhang *et al.*, 1998, *Nat. Genet.* 20(2): 123-28; incorporated herein by reference in its entirety). Homologous recombination can be performed between a PCR fragment flanked by short homology arms and an endogenous intact recipient such as a BAC. Using this method, homologous recombination is not limited by the disposition of restriction endonuclease cleavage sites or the size of the target DNA. A BAC can be modified in its host strain using a plasmid, *e.g.*, pBAD-αβγ, in which recE and recT have been replaced by their respective functional counterparts of phage lambda (Muyrers *et al.*, 1999, *Nucleic Acids Res.* 27(6): 1555-57). Preferably, a BAC is modified by recombination with a PCR product containing homology arms ranging from 27-60 bp. In a specific embodiment, homology arms are 50 bp in length.

In another embodiment, a transgene is inserted into a yeast artificial chromosome (YAC) (Burke *et al.*, 1987, *Science* 236: 806-12; and Peterson *et al.*, 1997, *Trends Genet.* 13:61, both of which are incorporated by reference herein in their entireties).

In other embodiments, the transgene is inserted into another vector developed for the cloning of large segments of genomic DNA, such as a cosmid or bacteriophage P1 (Sternberg *et al.*, 1990, *Proc. Natl. Acad. Sci. USA* 87: 103-07). The approximate maximum insert size is 30-35 kb for cosmids and 100 kb for bacteriophage P1. In another embodiment, the transgene is inserted into a P-1 derived artificial chromosome (PAC) (Mejia *et al.*, 1997, *Genome Res* 7:179-186). The maximum insert size is 300 kb.

Vectors containing the appropriate heterologous sequences may be identified by any method well known in the art, for example, by sequencing, restriction mapping, hybridization, PCR amplification, etc.

The vectors of the invention comprise one or more nucleotide sequences encoding a heterologous protein desired to be expressed in the transgenic avian, as well as regulatory elements such as promoters, enhancers, MARs, IRES's and other translation control elements, transcriptional termination elements, polyadenylation sequences, etc, as discussed *infra*. In particular embodiments, the vector of the invention contains at least two

nucleotide sequences coding for heterologous proteins, for example, but not limited to, the heavy and light chains of an immunoglobulin.

In a preferred embodiment, the nucleotide sequence encoding the heterologous protein is inserted into all or a significant portion of a nucleic acid containing the genomic sequence of an endogenous avian gene, preferably an avian gene that is expressed in the magnum, *e.g.*, ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin, etc. For example, the heterologous gene sequence may be inserted into or replace a portion of the 3' untranslated region (UTR) or 5' untranslated region (UTR) or an intron sequence of the endogenous gene genomic sequence. Preferably, the heterologous gene coding sequence has its own IRES. For descriptions of IRES's, *see, e.g.,* Jackson *et al.,* 1990, *Trends Biochem Sci.* 15(12):477-83; Jang *et al.*, 1988, *J. Virol.* 62(8):2636-43; Jang *et al.*, 1990, *Enzyme* 44(1-4):292-309; and Martinez-Salas, 1999, *Curr. Opin. Biotechnol.* 10(5):458-64; Palmenberg *et al.*, United States Patent No. 4,937,190, which are incorporated by reference herein in their entireties. In another embodiment, the heterologous protein coding sequence is inserted at the 3' end of the endogenous gene coding sequence. In another preferred embodiment, the heterologous gene coding sequences are inserted using 5' direct fusion wherein the heterologous gene coding sequences are inserted in-frame adjacent to the initial ATG sequence (or adjacent the nucleotide sequence encoding the first two, three, four, five, six, seven or eight amino acids) of the endogenous gene or replacing some or all of the sequence of the endogenous gene coding sequence. In yet another specific embodiment, the heterologous gene coding sequence is inserted into a separate cistron in the 5' region of the endogenous gene genomic sequence and has an independent IRES sequence.

The present invention further relates to nucleic acid vectors (preferably, not derived from eukaryotic viruses, except, in certain embodiments, for eukaryotic viral promoters and/or enhancers) and transgenes inserted therein that incorporate multiple polypeptide-encoding regions, wherein a first polypeptide-encoding region is operatively linked to a transcription promoter and a second polypeptide-encoding region is operatively linked to an IRES. For example, the vector may contain coding sequences for two different heterologous proteins (*e.g.*, the heavy and light chains of an immunoglobulin) or the coding sequences for all or a significant part of the genomic sequence for the gene from which the promoter driving expression of the transgene is derived, and the heterologous protein desired to be expressed (*e.g.*, a construct containing the genomic coding sequences, including introns, of the avian lysozyme gene when the avian lysozyme promoter is used to drive expression of the transgene, an IRES, and the coding sequence for the heterologous

protein desired to be expressed downstream (*i.e.*, 3' on the RNA transcript of the IRES). Thus, in certain embodiments, the nucleic acid encoding the heterologous protein is introduced into the 5' untranslated or 3' untranslated regions of an endogenous gene, such as but not limited to, ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin, with an IRES sequence directing translation of the heterologous sequence.

Such nucleic acid constructs, when inserted into the genome of a bird and expressed therein, will generate individual polypeptides that may be post-translationally modified, for example, glycosylated or, in certain embodiments, form complexes, such as heterodimers with each other in the white of the avian egg. Alternatively, the expressed polypeptides may be isolated from an avian egg and combined *in vitro*, or expressed in a non-reproductive tissue such as serum. In other embodiments, for example, but not limited to, when expression of both heavy and light chains of an antibody is desired, two separate constructs, each containing a coding sequence for one of the heterologous proteins operably linked to a promoter (either the same or different promoters), are introduced by microinjection into cytoplasm of one or more embryonic cells and transgenic avians harboring both transgenes in their genomes and expressing both heterologous proteins are identified. Alternatively, two transgenic avians each containing one of the two heterologous proteins (*e.g.*, one transgenic avian having a transgene encoding the light chain of an antibody and a second transgenic avian having a transgene encoding the heavy chain of the antibody) can be bred to obtain an avian containing both transgenes in its germline and expressing both transgene encoded proteins, preferably in eggs.

Recombinant expression vectors can be designed for the expression of the encoded proteins in eukaryotic cells. Useful vectors may comprise constitutive or inducible promoters to direct expression of either fusion or non-fusion proteins. With fusion vectors, a number of amino acids are usually added to the expressed target gene sequence such as, but not limited to, a protein sequence for thioredoxin, a polyhistidine, or any other amino acid sequence that facilitates purification of the expressed protein. A proteolytic cleavage site may further be introduced at a site between the target recombinant protein and the fusion sequence. Additionally, a region of amino acids such as a polymeric histidine region may be introduced to allow binding of the fusion protein to metallic ions such as nickel bonded to a solid support, and thereby allow purification of the fusion protein. Once the fusion protein has been purified, the cleavage site allows the target recombinant protein to be separated from the fusion sequence. Enzymes suitable for use in cleaving the proteolytic cleavage site include, but are not limited to, Factor Xa and thrombin. Fusion expression vectors that may be useful in the present invention include pGex (AMRAD® Corp.,

Melbourne, Australia), pRIT5 (PHARMACIA®, Piscataway, NJ) and pMAL (NEW ENGLAND BIOLABS®, Beverly, MA), fusing glutathione S-transferase, protein A, or maltose E binding protein, respectively, to the target recombinant protein.

Once a promoter and a nucleic acid encoding a heterologous protein of the present invention have been cloned into a vector system, it is ready to be incorporated into a host cell. Such incorporation can be carried out by the various forms of transformation noted above, depending upon the vector/host cell system. It is contemplated that the incorporation of the DNA of the present invention into a recipient cell may be by any suitable method such as, but not limited to, viral transfer, electroporation, gene gun insertion, sperm-mediated transfer to an ovum, microinjection and the like. Suitable host cells include, but are not limited to, bacteria, virus, yeast, mammalian cells, and the like. In particular, the present invention contemplates the use of recipient avian cells, such as chicken cells or quail cells.

Another aspect of the present invention, therefore, is a method of expressing a heterologous polypeptide in a eukaryotic cell by transfecting an avian cell with a recombinant DNA comprising an avian tissue-specific promoter operably linked to a nucleic acid insert encoding a polypeptide and, optionally, a polyadenylation signal sequence, and culturing the transfected cell in a medium suitable for expression of the heterologous polypeptide under the control of the avian lysozyme gene expression control region.

Yet another aspect of the present invention is a eukaryotic cell transformed with an expression vector according to the present invention and described above. In one embodiment of the present invention, the transformed cell is a chicken oviduct cell and the nucleic acid insert comprises the chicken lysozyme gene expression control region, a nucleic acid insert encoding a human interferon α2b and codon optimized for expression in an avian cell, and an SV40 polyadenylation sequence.

In another embodiment, the transformed cell is a quail oviduct cell and the nucleic acid insert comprises the artificial avian promoter construct MDOT (SEQ ID NO.:11) operably linked to an interferon-encoding sequence, as described in Example 23 below.

In yet another embodiment of the present invention, a quail oviduct cell is transfected with the nucleic acid insert comprising the MDOT artificial promoter construct operably linked to an erythropoietin (EPO)-encoding nucleic acid, wherein the transfected quail produces heterologous erythropoietin.

5.2.1 PROMOTERS

The vectors of the invention contain promoters that function in avian cells, preferably, that are tissue-specific and, in preferred embodiments, direct expression in the magnum or serum or other tissue such that expressed proteins are deposited in eggs, more preferably, that are specific for expression in the magnum. Alternatively, the promoter directs expression of the protein in the serum of the transgenic avian. Introduction of the vectors of the invention, preferably, generate transgenics that express the heterologous protein in tubular gland cells where it is secreted into the oviduct lumen and deposited, *e.g.*, into the white of an egg. In preferred embodiments, the promoter directs a level of expression of the heterologous protein in the egg white of eggs laid by G0 and/or G1 chicks and/or their progeny that is greater than 5 μg, 10 μg, 50 μg, 100 μg, 250 μg, 500 μg or 750 μg, more preferably greater than 1 mg, 2 mg, 5 mg, 10 mg, 20 mg, 50 mg, 100 mg, 200 mg, 500 mg, 700 mg, 1 gram, 2 grams, 3 grams, 4 grams or 5 grams. Such levels of expression can be obtained using the promoters of the invention.

In preferred embodiments, the promoters of the invention are derived from genes that express proteins present in significant levels in the egg white and/or the serum. For example, the promoter comprises regions of an ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin or ovomucin promoter or any other promoter that directs expression of a gene in an avian, particularly in a specific tissue of interest, such as the magnum or in the serum. Alternatively, the promoter used in the expression vector may be derived from that of the *lysozyme* gene that is expressed in both the oviduct and macrophages. Portions of two or more of these, and other promoters that function in avians, may be combined to produce effective synthetic promoter.

The promoter may optionally be a segment of the *ovalbumin* promoter region that is sufficiently large to direct expression of the coding sequence in the tubular gland cells. Other exemplary promoters include the promoter regions of the *ovalbumin*, *lysozyme*, *ovomucoid*, *conalbumin*, ovotransferrin or *ovomucin* genes (for example, but not limited to, as disclosed in co-pending United States Patent Application Nos. 09/922,549, filed August 3, 2001 and 10/114,739, filed April 1, 2002, both entitled "Avian Lysozyme Promoter", by Rapp, and United States Patent Application No. 09/998,716, filed November 30, 2001, entitled "Ovomucoid Promoter and Methods of Use," by Harvey *et al.*, all of which are incorporated by reference herein in their entireties). Alternatively, the promoter may be a promoter that is largely, but not entirely, specific to the magnum, such as the *lysozyme* promoter. Other suitable promoters may be artificial constructs such as a combination of nucleic acid regions derived from at least two avian gene promoters. One such embodiment

of the present invention is the MDOT construct (SEQ ID NO: 11) comprising regions derived from the chicken ovomucin and ovotransferrin promoters, including but not limited to promoters altered, *e.g.*, to increase expression, and inducible promoters, *e.g.*, the tet^r system.

The *ovalbumin* gene encodes a 45 kD protein that is also specifically expressed in the tubular gland cells of the magnum of the oviduct (Beato, 1989, *Cell* 56:335-344). Ovalbumin is the most abundant egg white protein, comprising over 50 percent of the total protein produced by the tubular gland cells, or about 4 grams of protein per large Grade A egg (Gilbert, "Egg albumen and its formation" in *Physiology and Biochemistry of the Domestic Fowl*, Bell and Freeman, eds., Academic Press, London, New York, pp. 1291-1329). The *ovalbumin* gene and over 20 kb of each flanking region have been cloned and analyzed (Lai *et al.*, 1978, *Proc. Natl. Acad. Sci. USA* 75:2205-2209; Gannon *et al.*, 1979, *Nature* 278:428-424; Roop *et al.*,1980, *Cell* 19:63-68; and Royal *et al.*,1975, *Nature* 279:125-132).

The *ovalbumin* gene responds to steroid hormones such as estrogen, glucocorticoids, and progesterone, which induce the accumulation of about 70,000 *ovalbumin* mRNA transcripts per tubular gland cell in immature chicks and 100,000 *ovalbumin* mRNA transcripts per tubular gland cell in the mature laying hen (Palmiter, 1973, *J. Biol. Chem.* 248:8260-8270; Palmiter, 1975, *Cell* 4:189-197). The 5' flanking region contains four DNAse I-hypersensitive sites centered at -0.25, -0.8, -3.2, and -6.0 kb from the transcription start site. These sites are called HS-I, -II, -III, and -IV, respectively. Promoters of the invention may contain one, all, or a combination of HS-I, HS-II, HS-III and HS0IV. Hypersensitivity of HS-II and -III are estrogen-induced, supporting a role for these regions in hormone-induction of *ovalbumin* gene expression.

HS-I and HS-II are both required for steroid induction of *ovalbumin* gene transcription, and a 1.4 kb portion of the 5' region that includes these elements is sufficient to drive steroid-dependent *ovalbumin* expression in explanted tubular gland cells (Sanders and McKnight, 1988, *Biochemistry* 27: 6550-6557). HS-I is termed the negative-response element ("NRE") because it contains several negative regulatory elements which repress *ovalbumin* expression in the absence of hormone (Haekers *et al.*, 1995, *Mol. Endo.* 9:1113-1126). Protein factors bind these elements, including some factors only found in oviduct nuclei suggesting a role in tissue-specific expression. HS-II is termed the steroid-dependent response element ("SDRE") because it is required to promote steroid induction of transcription. It binds a protein or protein complex known as Chirp-I. Chirp-I is induced by estrogen and turns over rapidly in the presence of cyclohexamide (Dean *et al.*, 1996, *Mol.*

Cell. Biol. 16:2015-2024). Experiments using an explanted tubular gland cell culture system defined an additional set of factors that bind SDRE in a steroid-dependent manner, including a NFκB-like factor (Nordstrom *et al.*, 1993, *J. Biol. Chem.* 268:13193-13202; Schweers and Sanders, 1991, *J. Biol. Chem.* 266: 10490-10497).

Less is known about the function of HS-III and HS-IV. HS-III contains a functional estrogen response element, and confers estrogen inducibility to either the *ovalbumin* proximal promoter or a heterologous promoter when co-transfected into HeLa cells with an estrogen receptor cDNA. These data imply that HS-III may play a functional role in the overall regulation of the *ovalbumin* gene. Little is known about the function of HS-IV, except that it does not contain a functional estrogen-response element (Kato *et al.*, 1992, *Cell* 68: 731-742).

In an alternative embodiment of the invention, transgenes containing constitutive promoters are used, but the transgenes are engineered so that expression of the transgene effectively becomes magnum-specific. Thus, a method for producing an exogenous protein in an avian oviduct provided by the present invention involves generating a transgenic avian having two transgenes in its tubular gland cells. One transgene comprises a first coding sequence operably linked to a constitutive promoter. The second transgene comprises a second coding sequence that is operably linked to a magnum-specific promoter, where expression of the first coding sequence is either directly or indirectly dependent upon the cellular presence of the protein expressed by the second coding sequence.

Additional promoters useful in the present invention include inducible promoters, such as the tet operator and the metallothionein promoter which can be induced by treatment with tetracycline and zinc ions, respectively (Gossen *et al.*, 1992, *Proc. Natl. Acad. Sci.* 89: 5547-5551 and Walden *et al.*, 1987, *Gene* 61: 317-327; incorporated herein by reference in their entireties).

5.2.1.1 CHICKEN LYSOZYME GENE EXPRESSION CONTROL REGION NUCLEIC ACID SEQUENCES

The chicken lysozyme gene is highly expressed in the myeloid lineage of hematopoietic cells, and in the tubular glands of the mature hen oviduct (Hauser *et al.*, 1981, *Hematol. and Blood Transfusion* 26: 175-178; Schutz *et al.*, 1978, Cold Spring Harbor Symp. Quart. Biol. 42: 617-624) and is therefore a suitable candidate for an efficient promoter for heterologous protein production in transgenic animals. The regulatory region of the lysozyme locus extends over at least 12 kb of DNA 5′ upstream of the transcription start site, and comprises a number of elements that have been individually isolated and

characterized. The known elements include three enhancer sequences at about -6.1 kb, -3.9 kb, and -2.7 kb (Grewal *et al.*, 1992, *Mol. Cell Biol.* 12: 2339-2350; Bonifer *et al.*, 1996, *J. Mol. Med.* 74: 663-671), a hormone responsive element (Hecht *et al.*, 1988, *E.M.B.O.J.* 7: 2063-2073), a silencer element and a complex proximal promoter. The constituent elements of the lysozyme gene expression control region are identifiable as DNAase 1 hypersensitive chromatin sites (DHS). They may be differentially exposed to nuclease digestion depending upon the differentiation stage of the cell. For example, in the multipotent progenitor stage of myelomoncytic cell development, or in erythroblasts, the silencer element is a DHS. At the myeloblast stage, a transcription enchancer located -6.1 kb upstream from the gene transcription start site is a DHS, while at the later monocytic stage another enhancer, at -2.7 kb becomes DNAase sensitive (Huber *et al.*, 1995, *DNA and Cell Biol.* 14: 397-402).

This invention also envisions the use of promoters other than the lysozyme promoter, including but not limited to, a cytomegalovirus promoter, an ovomucoid, conalbumin or ovotransferrin promoter or any other promoter that directs expression of a gene in an avian, particularly in a specific tissue of interest, such as the magnum.

Another aspect of the methods of the present invention is the use of combinational promoters comprising an artificial nucleic acid construct having at least two regions wherein the regions are derived from at least two gene promoters, including but not limited to a lysozyme, ovomucoid, conalbumin or ovotransferrin promoter. In one embodiment of the present invention, the promoter may comprise a region of an avian ovomucoid promoter and a region of an avian oxotransferrin promoter, thereby generating the MDOT avian artificial promoter construct. The avian MDOT promoter construct of the present invention has the nucleic acid sequence SEQ ID NO: 11 and is illustrated in Figure 7. This promoter is useful for allowing expression of a heterologous protein in chicken oviduct cells and may be operably linked to any nucleic acid encoding a heterologous polypeptide of interest including, for example, a cytokine, growth hormone, growth factor, enzyme, structural protein or the like.

5.2.2 MATRIX ATTACHMENT REGIONS

In preferred embodiments of the invention, the vectors contain matrix attachment regions (MARs) that preferably flank the transgene sequences to reduce position effects on expression when integrated into the avian genome. In fact, 5' MARs and 3' MARs (also referred to as "scaffold attachment regions" or SARs) have been identified in the outer boundaries of the chicken lysozyme locus (Phi-Van *et al.*, 1988, *E.M.B.O.J.* 7: 655-664;

Phi-Van, L. and Stratling, W.H., 1996, *Biochem.* 35: 10735-10742). Deletion of a 1.32 kb or a 1.45 kb halves region, each comprising half of a 5' MAR, reduces positional variation in the level of transgene expression (Phi-Van and Stratling, *supra*).

The 5' matrix-associated region (5' MAR), located about –11.7 kb upstream of the chicken lysozyme transcription start site, can increase the level of gene expression by limiting the positional effects exerted against a transgene (Phi-Van *et al.*, 1988, *supra*). At least one other MAR is located 3' downstream of the protein encoding region. Although MAR nucleic acid sequences are conserved, little cross-hybridization is seen, indicating significant overall sequence variation. However, MARs of different species can interact with the nucleomatrices of heterologous species, to the extent that the chicken lysozyme MAR can associate with the plant tobacco nucleomatrix as well as that of the chicken oviduct cells (Mlynarona *et al.*, 1994, *Cell* 6: 417-426; von Kries *et al.*, 1990, *Nucleic Acids Res.* 18: 3881-3885).

Gene expression must be considered not only from the perspective of cis-regulatory elements associated with a gene, and their interactions with trans-acting elements, but also with regard to the genetic environment in which they are located. Chromosomal positioning effects (CPEs), therefore, are the variations in levels of transgene expression associated with different locations of the transgene within the recipient genome. An important factor governing CPE upon the level of transgene expression is the chromatin structure around a transgene, and how it cooperates with the cis-regulatory elements. The cis-elements of the lysozyme locus are confined within a single chromatin domain (Bonifer *et al.*, 1996, *supra*; Sippel *et al.*, pgs. 133-147 in Eckstein F. & Lilley D.M.J. (eds), "Nucleic Acids and Molecular Biology", Vol. 3, 1989, Springer.

The lysozyme promoter region of chicken is active when transfected into mouse fibroblast cells and linked to a reporter gene such as the bacterial chloramphenicol acetyltransferase (CAT) gene. The promoter element is also effective when transiently transfected into chicken promacrophage cells. In each case, however, the presence of a 5' MAR element increased positional independency of the level of transcription (Stief *et al.*, 1989, *Nature* 341: 343-345; Sippel *et al.*, pgs. 257 – 265 in Houdebine L.M. (ed), "Transgenic Animals: Generation and Use").

The ability to direct the insertion of a transgene into a site in the genome of an animal where the positional effect is limited offers predictability of results during the development of a desired transgenic animal, and increased yields of the expressed product. Sippel and Steif disclose, in U.S. Patent No. 5,731,178, which is incorporated by reference herein in its entirety, methods to increase the expression of genes introduced into eukaryotic

cells by flanking a transcription unit with scaffold attachment elements, in particular the 5′ MAR isolated from the chicken lysozyme gene. The transcription unit disclosed by Sippel and Steif was an artificial construct that combined only the -6.1 kb enhancer element and the proximal promoter element (base position -579 to +15) from the lysozyme gene. Other promoter associated elements were not included. However, although individual cis-regulatory elements have been isolated and sequenced, together with short regions flanking DNA, the entire nucleic acid sequence comprising the functional 5′ upstream region of the lysozyme gene has not been determined in its entirety and therefore not employed as a functional promoter to allow expression of a heterologous transgene.

Accordingly, vectors of the invention comprise MARs, preferably both 5' and 3' MARs that flank the transgene, including the heterologous protein coding sequences and the regulatory sequences.

5.2.3 CODON-OPTIMIZED GENE EXPRESSION

Another aspect of the present invention provides nucleic acid sequences encoding heterologous polypeptides that are codon-optimized for expression in avian cells, and derivatives and fragments thereof. When a heterologous nucleic acid is to be delivered to a recipient cell for expression therein, the sequence of the nucleic acid sequence may be modified so that the codons are optimized for the codon usage of the recipient species. For example, if the heterologous nucleic acid is transfected into a recipient chicken cell, the sequence of the expressed nucleic acid insert is optimized for chicken codon usage. This may be determined from the codon usage of at least one, and preferably more than one, protein expressed in a chicken cell. For example, the codon usage may be determined from the nucleic acid sequences encoding the proteins ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin of chicken. Briefly, the DNA sequence for the target protein may be optimized using the BACKTRANSLATE® program of the Wisconsin Package, version 9.1 (Genetics Computer Group, Inc., Madison, WI) with a codon usage table compiled from the chicken (*Gallus gallus*) ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin proteins. The template and primer oligonucleotides are then amplified, by any means known in the art, including but not limited to PCR with *Pfu* polymerase (STRATAGENE®, La Jolla CA).

In one exemplary embodiment of a heterologous nucleic acid for use by the methods of the present invention, a nucleic acid insert encoding the human interferon α2b polypeptide optimized for codon-usage by the chicken is microinjected into the cytoplasm

of a stage 1 embryo. Optimization of the sequence for codon usage is useful in elevating the level of translation in avian eggs.

It is contemplated to be within the scope of the present invention for any nucleic acid encoding a polypeptide to be optimized for expression in avian cells. It is further contemplated that the codon usage may be optimized for a particular avian species used as a source of the host cells. In one embodiment of the present invention, the heterologous polypeptide is encoded using the codon-usage of a chicken.

5.2.4 SPECIFIC VECTORS OF THE INVENTION

In a preferred embodiment, a transgene of the invention comprises a chicken, or other avian, lysozyme control region sequence which directs expression of the coding sequence within the transgene. A series of PCR amplifications of template chicken genomic DNA are used to isolate the gene expression control region of the chicken lysozyme locus. Two amplification reactions used the PCR primer sets 5pLMAR2 (5'-TGCCGCCTTCTTTGATATTC-3') (SEQ ID NO: 1) and LE-6.1kbrev1 (5'-TTGGTGGTAAGGCCTTTTTG-3') (SEQ ID NO: 2) (Set 1) and lys-6.1 (5'-CTGGCAAGCTGTCAAAAACA-3') (SEQ ID NO: 3) and LysE1Rev (5'-CAGCTCACATCGTCCAAAGA-3') (SEQ ID NO: 4) (Set 2). The amplified PCR products were united as a contiguous isolated nucleic acid by a third PCR amplification step with the primers SEQ ID NOS: 1 and 4.

The isolated PCR-amplified product, comprising about 12 kb of the nucleic acid region 5' upstream of the native chicken lysozyme gene locus, was cloned into the plasmid pCMV-LysSPIFNMM. pCMV-LysSPIFNMM comprises a modified nucleic acid insert encoding a human interferon α2b sequence and an SV40 polyadenylation signal sequence (SEQ ID NO: 8) 3′ downstream of the interferon encoding nucleic acid. The sequence SEQ ID NO: 5 of the nucleic acid insert encoding human interferon α2b was in accordance with avian cell codon usage, as determined from the nucleotide sequences encoding chicken ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin.

The nucleic acid sequence (SEQ ID NO: 6) (GenBank Accession No. AF405538) of the insert in pAVIJCR-A115.93.1.2 is shown in Figures 1A-E. The modified human interferon α2b encoding nucleotide sequence SEQ ID NO: 5 (GenBank Accession No. AF405539) and the novel chicken lysozyme gene expression control region SEQ ID NO: 7 (GenBank Accession No. AF405540), shown in Figures 2 and 3A-E, respectively. A polyadenylation signal sequence that is suitable for operably linking to the polypeptide-

encoding nucleic acid insert is the SV40 signal sequence SEQ ID NO: 8, as shown in Figure 4.

The plasmid pAVIJCR-A115.93.1.2 was restriction digested with enzyme *Fse*I to isolate a 15.4 kb DNA containing the lysozyme 5' matrix attachment region (MAR) and the -12.0 kb lysozyme promoter during the expression of the interferon-encoding insert, as described in Example 17, below. Plasmid pIIIilys was restriction digested with *Mlu*I and *Xho*I to isolate an approximately 6 kb nucleic acids, comprising the 3' lysozyme domain, the sequence of which (SEQ ID NO: 9) is shown in Figures 5A-C. The 15.4 kb and 6 kb nucleic acids were ligated and the 21.4 kb nucleic acid comprising the nucleic acid sequence SEQ ID NO: 10 as shown in Figures 6A-J was transformed into recipient STBL4 cells.

The inclusion of the novel isolated avian lysozyme gene expression control region of the present invention upstream of a codon-optimized interferon-encoding sequence in pAVIJCR-A115.93.1.2 allowed expression of the interferon polypeptide in avian cells transfected by sperm-mediated transfection. The 3' lysozyme domain SEQ ID NO: 9, when operably linked downstream of a heterologous nucleic acid insert, also allows expression of the nucleic acid insert as described in Example 18, below. For example, the nucleic acid insert may encode a heterologous polypeptide such as the α2b interferon encoded by the sequence SEQ ID NO: 5.

It is further contemplated that any nucleic acid sequence encoding a polypeptide may be operably linked to the novel isolated avian lysozyme gene expression control region (SEQ ID NO: 7) and optionally operably linked to the 3' lysozyme domain SEQ ID NO. 9 so as to be expressed in a transfected avian cell. The plasmid construct pAVIJCR-A115.93.1.2 can be introduced into cultured quail oviduct cells by transfection. ELISA assays of the cultured media showed that the transfected cells synthesized a polypeptide detectable with anti-human interferon α2b antibodies.

The isolated chicken lysozyme gene expression control region (SEQ ID NO: 7) for use in the methods of the present invention comprises the nucleotide elements that are positioned 5′ upstream of the lysozyme-encoding region of the native chicken lysozyme locus and which are necessary for the regulated expression of a downstream polypeptide-encoding nucleic acid. While not wishing to be bound by any one theory, the inclusion of at least one 5′ MAR sequence of or reference element in the isolated control region may confer positional independence to a transfected gene operably linked to the novel lysozyme gene expression control region.

The isolated lysozyme gene expression control region (SEQ ID NO: 7) of the present invention is useful for reducing the chromosomal positional effect of a transgene operably

linked to the lysozyme gene expression control region and transfected into a recipient avian cell. By isolating a region of the avian genome extending from a point 5′ upstream of a 5′ MAR of the lysozyme locus to the junction between the signal peptide sequence and a polypeptide-encoding region, cis-regulatory elements are also included that may allow gene expression in a tissue-specific manner. The lysozyme promoter region of the present invention, therefore, will allow expression of an operably linked heterologous nucleic acid insert in a transfected avian cell such as, for example, an oviduct cell.

It is further contemplated that a recombinant DNA of the present invention may further comprise the chicken lysozyme 3' domain (SEQ. ID NO: 9) linked downstream of the nucleic acid insert encoding a heterologous polypeptide. The lysozyme 3' domain (SEQ ID NO: 9) includes a nucleic acid sequence encoding a 3' MAR domain that may cooperate with a 5' MAR to direct the insertion of the construct of the present invention into the chromosome of a transgenic avian, or may act independently of the 5' MAR.

Fragments of a nucleic acid encoding a portion of the subject lysozyme gene expression control region may also be useful as an autonomous gene regulatory element that may itself be operably linked to a polypeptide-encoding nucleic acid. Alternatively, the fragment may be combined with fragments derived from other gene promoters, such as an avian ovalbumin, vomucoid, ovotransferrin, conalbumin or ovomucin promoter, thereby generating novel promoters having new properties or a combination of properties. As used herein, a fragment of the nucleic acid encoding an active portion of a lysozyme gene expression control region refers to a nucleotide sequence having fewer nucleotides than the nucleotide sequence encoding the entire nucleic acid sequence of the lysozyme gene expression control region, but at least 200 nucleotides.

The present invention also contemplates the use of antisense nucleic acid molecules that are designed to be complementary to a coding strand of a nucleic acid (*i.e.*, complementary to an endogenous DNA or an mRNA sequence) or, alternatively, complimentary to a 5' or 3' untranslated region of the mRNA and therefore useful for regulating the expression of a gene by the lysozyme promoter.

Synthesized oligonucleotides can be produced in variable lengths when for example, non-naturally occurring polypeptide sequences are desired. The number of bases synthesized will depend upon a variety of factors, including the desired use for the probes or primers. Additionally, sense or anti-sense nucleic acids or oligonucleotides can be chemically synthesized using modified nucleotides to increase the biological stability of the molecule or of the binding complex formed between the anti-sense and sense nucleic acids. For example, acridine substituted nucleotides can be synthesized. Protocols for designing

isolated nucleotides, nucleotide probes, and/or nucleotide primers are well-known to those of ordinary skill, and can be purchased commercially from a variety of sources (*e.g.*, SIGMA GENOSYS®, The Woodlands, TX or The Great American Gene Co., Ramona, CA).

5.2.5 RECOMBINANT EXPRESSION VECTORS

A useful application of the novel promoters of the present invention, such as the avian lysozyme gene expression control region (SEQ ID NO: 7) or the MDOT promoter construct (SEQ ID NO: 11) is the possibility of increasing the amount of a heterologous protein present in a bird, especially a chicken, by gene transfer. In most instances, a heterologous polypeptide-encoding nucleic acid insert transferred into the recipient animal host will be operably linked with a gene expression control region to allow the cell to initiate and continue production of the genetic product protein. A recombinant DNA molecule of the present invention can be transferred into the extra-chromosomal or genomic DNA of the host.

Expression of a foreign gene in an avian cell permits partial or complete post-translational modification such as, but not only, glycosylation, and/or the formation of the relevant inter- or intra-chain disulfide bonds. Examples of vectors useful for expression in the chicken *Gallus gallus* include pYepSecl (Baldari *et al.*, 1987, *E.M.B.O.J.*, 6: 229-234; incorporated herein by reference in its entirety) and pYES2 (INVITROGEN® Corp., San Diego, CA).

The present invention contemplates that the injected cell may transiently contain the injected DNA, whereby the recombinant DNA or expression vector may not be integrated into the genomic nucleic acid. It is further contemplated that the injected recombinant DNA or expression vector may be stably integrated into the genomic DNA of the recipient cell, thereby replicating with the cell so that each daughter cell receives a copy of the injected nucleic acid. It is still further contemplated for the scope of the present invention to include a transgenic animal producing a heterologous protein expressed from an injected nucleic acid according to the present invention.

Heterologous nucleic acid molecules can be delivered to oocytes using the sperm-mediated transfection methods of the present invention. The nucleic acid molecule may be inserted into a cell to which the nucleic acid molecule (or promoter coding region) is heterologous (*i.e.*, not normally present). Alternatively, the recombinant DNA molecule may be introduced into cells which normally contain the recombinant DNA molecule or the

particular coding region, as, for example, to correct a deficiency in the expression of a polypeptide, or where over-expression of the polypeptide is desired.

Another aspect of the present invention, therefore, is a method of expressing a heterologous polypeptide in an avian cell by transfecting the avian cell with a selected heterologous nucleic acid comprising an avian promoter operably linked to a nucleic acid insert encoding a polypeptide and, optionally, a polyadenylation signal sequence. The transfected cell, which may be an avian embryonic cell microinjected with a heterologous nucleic acid, will generate a transgenic embryo that after introduction into a recipient hen will be laid as a hard-shell egg and develop into a transgenic chick.

In another embodiment of the present invention, the nucleic acid insert comprises the chicken lysozyme gene expression control region, a nucleic acid insert encoding a human interferon α2b and codon optimized for expression in an avian cell, and a chicken 3' domain, *i.e.*, downstream enhancer elements.

In one embodiment of the present invention, the transgenic animal is an avian selected from a turkey, duck, goose, quail, pheasant, ratite, and ornamental bird or a feral bird. In another embodiment, the avian is a chicken and the heterologous polypeptide produced under the transcriptional control of the avian promoter is produced in the white of an egg. In yet another embodiment of the present invention, the heterologous polypeptide is produced in the serum of a bird.

5.3 HETEROLOGOUS PROTEINS PRODUCED BY TRANSGENIC AVIANS

Methods of the present invention, providing for the production of heterologous protein in the avian oviduct (or other tissue leading to deposition of the protein into the egg) and the production of eggs containing heterologous protein, involve providing a suitable vector coding for the heterologous protein and introducing the vector into oocytes by sperm-mediated transfection such that the vector is integrated into the genome of the resulting transgenic embryo. A subsequent step involves deriving a mature transgenic avian from the transgenic embryo produced in the previous steps by transferring the injected cell or cells into the infundibulum of a recipient hen; producing a hard shell egg from that hen; and allowing the egg to develop and hatch to produce a transgenic bird.

A transgenic avian so produced from transgenic embryonic cells is known as a founder. Such founders may be mosaic for the transgene (in certain embodiments, the founder has 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 90%, 100% of the cells containing the transgene. The invention further provides production of heterologous proteins in other

tissues of the transgenic avians. Some founders will carry the transgene in the tubular gland cells in the magnum of their oviducts. These birds will express the exogenous protein encoded by the transgene in their oviducts. If the exogenous protein contains the appropriate signal sequences, it will be secreted into the lumen of the oviduct and into the white of an egg.

Some founders are germ-line founders. A germ-line founder is a founder that carries the transgene in genetic material of its germ-line tissue, and may also carry the transgene in oviduct magnum tubular gland cells that express the exogenous protein. Therefore, in accordance with the invention, the transgenic bird may have tubular gland cells expressing the exogenous protein and the offspring of the transgenic bird will also have oviduct magnum tubular gland cells that express the exogenous protein. Alternatively, the offspring express a phenotype determined by expression of the exogenous gene in a specific tissue of the avian. In preferred embodiments, the heterologous proteins are produced from transgenic avians that were not (or the founder ancestors were not) using a eukaryotic viral vector, or a retroviral vector.

The present invention can be used to express, in large yields and at low cost, a wide range of desired proteins including those used as human and animal pharmaceuticals, diagnostics, and livestock feed additives. Proteins such as growth hormones, cytokines, structural proteins and enzymes, including human growth hormone, interferon, lysozyme, and β-casein, are examples of proteins that are desirably expressed in the oviduct and deposited in eggs according to the invention. Other possible proteins to be produced include, but are not limited to, albumin, α-1 antitrypsin, antithrombin III, collagen, factors VIII, IX, X (and the like), fibrinogen, hyaluronic acid, insulin, lactoferrin, protein C, erythropoietin (EPO), granulocyte colony-stimulating factor (G-CSF), granulocyte macrophage colony-stimulating factor (GM-CSF), tissue-type plasminogen activator (tPA), feed additive enzymes, somatotropin, and chymotrypsin Immunoglobulins and genetically engineered antibodies, including immunotoxins that bind to surface antigens on human tumor cells and destroy them, can also be expressed for use as pharmaceuticals or diagnostics. It is contemplated that immunoglobulin polypeptides expressed in avian cells following transfection by the methods of the present invention may include monomeric heavy and light chains, single-chain antibodies or multimeric immunoglobulins comprising variable heavy and light chain regions, *i.e.*, antigen-binding domains, or intact heavy and light immunoglobulin chains.

5.3.1 MULTIMERIC PROTEINS

The invention, in preferred embodiments, provides methods for producing multimeric proteins, preferably immunoglobulins, such as antibodies, and antigen binding fragments thereof.

In one embodiment of the present invention, the multimeric protein is an immunoglobulin, wherein the first and second heterologous polypeptides are an immunoglobulin heavy and light chains respectively. Illustrative examples of this and other aspects and embodiments of the present invention for the production of heterologous multimeric polypeptides in avian cells are fully disclosed in U.S. Patent Application No. 09/877,374, filed June 8, 2001, by Rapp, which is incorporated herein by reference in its entirety. In one embodiment of the present invention, therefore, the multimeric protein is an immunoglobulin wherein the first and second heterologous polypeptides are an immunoglobulin heavy and light chain respectively. Accordingly, the invention provides immunoglobulin and other multimeric proteins that have been produced by transgenic avians of the invention.

In the various embodiments of this aspect of the present invention, an immunoglobulin polypeptide encoded by the transcriptional unit of at least one expression vector may be an immunoglobulin heavy chain polypeptide comprising a variable region or a variant thereof, and may further comprise a D region, a J region, a C region, or a combination thereof. An immunoglobulin polypeptide encoded by the transcriptional unit of an expression vector may also be an immunoglobulin light chain polypeptide comprising a variable region or a variant thereof, and may further comprise a J region and a C region. It is also contemplated to be within the scope of the present invention for the immunoglobulin regions to be derived from the same animal species, or a mixture of species including, but not only, human, mouse, rat, rabbit and chicken. In preferred embodiments, the antibodies are human or humanized.

In other embodiments of the present invention, the immunoglobulin polypeptide encoded by the transcriptional unit of at least one expression vector comprises an immunoglobulin heavy chain variable region, an immunoglobulin light chain variable region, and a linker peptide thereby forming a single-chain antibody capable of selectively binding an antigen.

Another aspect of the present invention provides a method for the production in an avian of an heterologous protein capable of forming an antibody suitable for selectively binding an antigen comprising the step of producing a transgenic avian incorporating at least one transgene, wherein the transgene encodes at least one heterologous polypeptide

selected from an immunoglobulin heavy chain variable region, an immunoglobulin heavy chain comprising a variable region and a constant region, an immunoglobulin light chain variable region, an immunoglobulin light chain comprising a variable region and a constant region, and a single-chain antibody comprising two peptide-linked immunoglobulin variable regions. Preferably, the antibody is expressed such that it is deposited in the white of the developing eggs of the avian. The hard shell avian eggs thus produced can be harvested and the heterologous polypeptide capable of forming or which formed an antibody can be isolated from the harvested egg. It is also understood that the heterologous polypeptides may also be expressed under the transcriptional control of promoters that allow for release of the polypeptides into the serum of the transgenic animal. Exemplary promoters for non-tissue specific production of a heterologous protein are the CMV promoter and the RSV promoter.

In one embodiment of this method of the present invention, the transgene comprises a transcription unit encoding a first and a second immunoglobulin polypeptide operatively linked to a transcription promoter, a transcription terminator and, optionally, an internal ribosome entry site (IRES) (see, for example, U.S. Patent No. 4,937,190 to Palmenberg *et al.*, the contents of which is incorporated herein by reference in its entirety).

In an embodiment of this method of the present invention, the isolated heterologous protein is an antibody capable of selectively binding to an antigen. In this embodiment, the antibody may be generated within the serum of an avian or within the white of the avian egg by combining at least one immunoglobulin heavy chain variable region and at least one immunoglobulin light chain variable region, preferably cross-linked by at least one di-sulfide bridge. The combination of the two variable regions will generate a binding site capable of binding an antigen using methods for antibody reconstitution that are well known in the art.

It is, however, contemplated to be within the scope of the present invention for immunoglobulin heavy and light chains, or variants or derivatives thereof, to be expressed in separate transgenic avians, and therefore isolated from separate media including serum or eggs, each isolate comprising a single species of immunoglobulin polypeptide. The method may further comprise the step of combining a plurality of isolated heterologous immunoglobulin polypeptides, thereby producing an antibody capable of selectively binding to an antigen. In this embodiment, two individual transgenic avians may be generated wherein one transgenic produces serum or eggs having an immunoglobulin heavy chain variable region, or a polypeptide comprising such, expressed therein. A second transgenic animal, having a second transgene, produces serum or eggs having an immunoglobulin light

chain variable region, or a polypeptide comprising such, expressed therein. The polypeptides may be isolated from their respective sera and eggs and combined *in vitro* to generate a binding site capable of binding an antigen.

Examples of therapeutic antibodies that can be used in methods of the invention include but are not limited to HERCEPTIN® (Trastuzumab) (Genentech, CA) which is a humanized anti-HER2 monoclonal antibody for the treatment of patients with metastatic breast cancer; REOPRO® (abciximab) (Centocor) which is an anti-glycoprotein IIb/IIIa receptor on the platelets for the prevention of clot formation; ZENAPAX® (daclizumab) (Roche Pharmaceuticals, Switzerland) which is an immunosuppressive, humanized anti-CD25 monoclonal antibody for the prevention of acute renal allograft rejection; PANOREX™ which is a murine anti-17-IA cell surface antigen IgG2a antibody (Glaxo Wellcome/Centocor); BEC2 which is a murine anti-idiotype (GD3 epitope) IgG antibody (ImClone System); IMC-C225 which is a chimeric anti-EGFR IgG antibody (ImClone System); VITAXIN™ which is a humanized anti-αVβ3 integrin antibody (Applied Molecular Evolution/MedImmune); Campath 1H/LDP-03 which is a humanized anti CD52 IgG1 antibody (Leukosite); Smart M195 which is a humanized anti-CD33 IgG antibody (Protein Design Lab/Kanebo); RITUXAN™ which is a chimeric anti-CD20 IgG1 antibody (IDEC Pharm/Genentech, Roche/Zettyaku); LYMPHOCIDE™ which is a humanized anti-CD22 IgG antibody (Immunomedics); ICM3 is a humanized anti-ICAM3 antibody (ICOS Pharm); IDEC-114 is a primatied anti-CD80 antibody (IDEC Pharm/Mitsubishi); ZEVALIN™ is a radiolabelled murine anti-CD20 antibody (IDEC/Schering AG); IDEC-131 is a humanized anti-CD40L antibody (IDEC/Eisai); IDEC-151 is a primatized anti-CD4 antibody (IDEC); IDEC-152 is a primatized anti-CD23 antibody (IDEC/Seikagaku); SMART anti-CD3 is a humanized anti-CD3 IgG (Protein Design Lab); 5G1.1 is a humanized anti-complement factor 5 (C5) antibody (Alexion Pharm); D2E7 is a humanized anti-TNF-α antibody (CAT/BASF); CDP870 is a humanized anti-TNF-α Fab fragment (Celltech); IDEC-151 is a primatized anti-CD4 IgG1 antibody (IDEC Pharm/SmithKline Beecham); MDX-CD4 is a human anti-CD4 IgG antibody (Medarex/Eisai/Genmab); CDP571 is a humanized anti-TNF-α IgG4 antibody (Celltech); LDP-02 is a humanized anti-α4β7 antibody (LeukoSite/Genentech); OrthoClone OKT4A is a humanized anti-CD4 IgG antibody (Ortho Biotech); ANTOVA™ is a humanized anti-CD40L IgG antibody (Biogen); ANTEGREN™ is a humanized anti-VLA-4 IgG antibody (Elan); and CAT-152 is a human anti-TGF-β_2 antibody (Cambridge Ab Tech).

5.3.2 PROTEIN RECOVERY

The protein of the present invention may be produced in purified form by any known conventional technique. For example, chicken cells may be homogenized and centrifuged. The supernatant can then be subjected to sequential ammonium sulfate precipitation and heat treatment. The fraction containing the protein of the present invention is subjected to gel filtration in an appropriately sized dextran or polyacrylamide column to separate the proteins. If necessary, the protein fraction may be further purified by HPLC. In another embodiment, an affinity column is used, wherein the protein is expressed with a tag.

Accordingly, the invention provides proteins that are produced by transgenic avians of the invention. In a preferred embodiment, the protein is produced and isolated from an avian egg. In another embodiment, the protein is produced and isolated from avian serum.

5.4 PHARMACEUTICAL COMPOSITIONS

The present invention further provides pharmaceutical compositions, formulations, dosage units and methods of administration comprising the heterologous proteins produced by the transgenic avians using methods of the invneion. Preferably, compositions of the invention comprise a prophylactically or therapeutically effective amount of a the heterologous protein, and a pharmaceutically acceptable carrier.

The term "carrier" refers to a diluent, adjuvant, excipient, or vehicle with which a compound of the invention is administered. Such pharmaceutical vehicles can be liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. The pharmaceutical vehicles can be saline, gum acacia, gelatin, starch paste, talc, keratin, colloidal silica, urea, and the like. In addition, auxiliary, stabilizing, thickening, lubricating and coloring agents may be used. When administered to a patient, the compounds of the invention and pharmaceutically acceptable vehicles are preferably sterile. Water is a preferred vehicle when the compound of the invention is administered intravenously. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid vehicles, particularly for injectable solutions. Suitable pharmaceutical vehicles also include excipients such as starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propyleneglycol, water, ethanol and the like. The present compositions, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents.

The present compositions can take the form of solutions, suspensions, emulsion, tablets, pills, pellets, capsules, capsules containing liquids, powders, sustained-release

formulations, suppositories, emulsions, aerosols, sprays, suspensions, or any other form suitable for use. In one embodiment, the pharmaceutically acceptable vehicle is a capsule (see *e.g.*, U.S. Patent No. 5,698,155). Other examples of suitable pharmaceutical vehicles are described in "Remington: the Science and Practice of Pharmacy", 20th ed., by Mack Publishing Co. 2000.

In a preferred embodiment, the heterologous proteins are formulated in accordance with routine procedures as a pharmaceutical composition adapted for intravenous administration to human beings. Typically, compounds of the invention for intravenous administration are solutions in sterile isotonic aqueous buffer. Where necessary, the compositions may also include a solubilizing agent. Compositions for intravenous administration may optionally include a local anesthetic such as lignocaine to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the heterologous protein of the invention is to be administered by infusion, it can be dispensed, for example, with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the composition of the invention is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

Compositions for oral delivery may be in the form of tablets, lozenges, aqueous or oily suspensions, granules, powders, emulsions, capsules, syrups, or elixirs, for example. Orally administered compositions may contain one or more optional agents, for example, sweetening agents such as fructose, aspartame or saccharin; flavoring agents such as peppermint, oil of wintergreen, or cherry; coloring agents; and preserving agents, to provide a pharmaceutically palatable preparation. Moreover, where in tablet or pill form, the compositions may be coated to delay disintegration and absorption in the gastrointestinal tract thereby providing a sustained action over an extended period of time. Selectively permeable membranes surrounding an osmotically active driving compound are also suitable for orally administered compounds of the invention. In these later platforms, fluid from the environment surrounding the capsule is imbibed by the driving compound, which swells to displace the agent or agent composition through an aperture. These delivery platforms can provide an essentially zero order delivery profile as opposed to the spiked profiles of immediate release formulations. A time delay material such as glycerol monostearate or glycerol stearate may also be used. Oral compositions can include standard

vehicles such as mannitol, lactose, starch, magnesium stearate, sodium saccharin, cellulose, magnesium carbonate, *etc*. Such vehicles are preferably of pharmaceutical grade.

Further, the effect of the heterologous proteins may be delayed or prolonged by proper formulation. For example, a slowly soluble pellet of the compound may be prepared and incorporated in a tablet or capsule. The technique may be improved by making pellets of several different dissolution rates and filling capsules with a mixture of the pellets. Tablets or capsules may be coated with a film which resists dissolution for a predictable period of time. Even the parenteral preparations may be made long-acting, by dissolving or suspending the compound in oily or emulsified vehicles which allow it to disperse only slowly in the serum.

5.5 TRANSGENIC AVIANS

Another aspect of the present invention concerns transgenic avians, preferably chicken or quail, produced by methods of the invention described in section 5.1 *infra*, preferably by introducing a nucleic acid comprising a transgene into an avian oocyte by the sperm-mediated transfection methods of the present invention. In one embodiment, a heterologous nucleic acid introduced to an avian oocyte by sperm-mediated transfection, resulting in a transgenic embryo which is then allowed to develop, preferably, transferred into the reproductive tract of a recipient hen where it is encapsulated by natural egg white proteins and a natural egg shell, then it is incubated and hatched to produce a transgenic chick. The heterologous polypeptide or polypeptides encoded by the transgenic heterologous nucleic acid may be secreted into the oviduct lumen of the mature transgenic chicken and deposited as a constituent component of egg white. The resulting transgenic avian chick (*i.e*, the G0) will carry one or more desired transgene(s) some or all of its cells, preferably in its germ line. These G0 transgenic avians can be bred using methods well known in the art to generate second generation (*i.e.*, G1s) transgenic avians that carry the transgene, *i.e.*, achieve germline transmission of the transgene. In preferred embodiments, the methods of the invention result in germline transmission, *i.e.*, percentage of G0s that transmit the transgene to progeny (G1s), that is greater than 5%, preferably, greater than 10%, 20%, 30%, 40%, and, most preferably, greater than 50%, 60%, 70%, 80%, 90% or even 100%. In other embodiments, the efficiency of transgenesis (i.e., number of G0s containing the transgene) is greater than 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% or 99%.

The egg can be harvested after laying and before hatching of a chick, or further incubated to generate a cloned chick, optionally genetically modified. The cloned chick

may carry a transgene in all or most of its cells. After maturation, the transgenic avian may lay eggs that contain one or more desired heterologous protein(s).

The cloned chick may also be a knock-in chick expressing an alternative phenotype or capable of laying eggs having an heterologous protein therein. The reconstructed egg may also be cultured to term using the *ex ovo* method described by Perry *et al.* (*supra*).

Following maturation, the transgenic avian and/or transgenic progeny thereof, may lay eggs containing one or more desired heterologous protein(s) expressed therein and that can be easily harvested therefrom. The G1 chicks, when sexually mature, can then be bred to produce progeny that are homozygous or heterozygous for the transgene.

A transgenic avian of the invention may contain at least one transgene, at least two transgenes, at least 3 transgenes, at least 4 transgenes, at least 5 transgenes, and preferably, though optionally, may express the subject nucleic acid encoding a polypeptide in one or more cells in the animal, such as the oviduct cells of the chicken. In embodiments of the present invention, the expression of the transgene may be restricted to specific subsets of cells, tissues, or developmental stages utilizing, for example, cis-acting sequences that control expression in the desired pattern. Toward this end, it is contemplated that tissue-specific regulatory sequences, or tissue-specific promoters, and conditional regulatory sequences may be used to control expression of the transgene in certain spatial patterns. Moreover, temporal patterns of expression can be provided by, for example, conditional recombination systems or prokaryotic transcriptional regulatory sequences. The inclusion of a 5' MAR region, and optionally the 3' MAR on either end of the sequence, in the expression cassettes suitable for use in the methods of the present invention may allow the heterologous expression unit to escape the chromosomal positional effect (CPE) and therefore be expressed at a more uniform level in transgenic tissues that received the transgene by a route other than through germ line cells.

The transgenes may, in certain embodiments, be expressed conditionally, *e.g.*, the heterologous protein coding sequence is under the control of an inducible promoter, such as a prokaryotic promoter or operator that requires a prokaryotic inducer protein to be activated. Operators present in prokaryotic cells have been extensively characterized *in vivo* and *in vitro* and can be readily manipulated to place them in any position upstream from or within a gene by standard techniques. Such operators comprise promoter regions and regions that specifically bind proteins such as activators and repressors. One example is the operator region of the *lexA* gene of *E. coli* to which the LexA polypeptide binds. Other exemplary prokaryotic regulatory sequences and the corresponding trans-activating prokaryotic proteins are disclosed by Brent and Ptashne in U.S. Patent No. 4,833,080 (the

contents of which is herein incorporated by reference in its entirety). Transgenic animals can be created which harbor the subject transgene under transcriptional control of a prokaryotic sequence or other activator sequence that is not appreciably activated by avian proteins. Breeding of this transgenic animal with another animal that is transgenic for the corresponding trans-activator can be used to activate of the expression of the transgene. . Moreover, expression of the conditional transgenes can also be induced by gene therapy-like methods wherein a gene encoding the trans-activating protein, *e.g.*, a recombinase or a prokaryotic protein, is delivered to the tissue and caused to be expressed, such as in a cell-type specific manner.

Transactivators in these inducible or repressible transcriptional regulation systems are designed to interact specifically with sequences engineered into the transgene. Such systems include those regulated by tetracycline ("tet systems"), interferon, estrogen, ecdysone, Lac operator, progesterone antagonist RU486, and rapamycin (FK506) with tet systems being particularly preferred (see, *e.g.*, Gingrich and Roder, 1998, *Annu. Rev. Neurosci.* 21: 377-405; incorporated herein by reference in its entirety). These drugs or hormones (or their analogs) act on modular transactivators composed of natural or mutant ligand-binding domains and intrinsic or extrinsic DNA binding and transcriptional activation domains. In certain embodiments, expression of the heterologous peptidecan be regulated by varying the concentration of the drug or hormone in medium *in vitro* or in the diet of the transgenic animal *in vivo*.

In a preferred embodiment, the control elements of the tetracycline-resistance operon of *E. coli* is used as an inducible or repressible transactivator or transcriptional regulation system ("tet system") for conditional expression of the transgene. A tetracycline-controlled transactivator can require either the presence or absence of the antibiotic tetracycline, or one of its derivatives, *e.g.*, doxycycline (dox), for binding to the tet operator of the tet system, and thus for the activation of the tet system promoter (Ptet).

In a specific embodiment, a tetracycline-repressed regulatable system (TrRS) is used (Agha-Mohammadi and Lotze, 2000, *J. Clin. Invest.* 105(9): 1177-83; Shockett *et al.*, 1995, *Proc. Natl. Acad. Sci. USA* 92: 6522-26; and Gossen and Bujard, 1992, *Proc. Natl. Acad. Sci. USA* 89: 5547-51; incorporated herein by reference in their entireties).

In another embodiment, a reverse tetracycline-controlled transactivator, *e.g.*, rtTA2 S-M2, is used. rtTA2 S-M2 transactivator has reduced basal activity in the absence doxycycline, increased stability in eukaryotic cells, and increased doxycycline sensitivity (Urlinger *et al.*, 2000, *Proc. Natl. Acad. Sci. USA* 97(14): 7963-68; incorporated herein by reference in its entirety). In another embodiment, the tet-repressible system described by

Wells *et al.* (1999, *Transgenic Res.* 8(5): 371-81; incorporated herein by reference in its entirety) is used. In one aspect of the embodiment, a single plasmid Tet-repressible system is used.. In another embodiment, the GAL4-UAS system (Ornitz *et al.*, 1991, *Proc. Natl. Acad. Sci. USA* 88:698-702; Rowitch *et al.*, 1999, *J. Neuroscience* 19(20):8954–8965; Wang *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96:8483-8488; Lewandoski, 2001, *Nature Reviews (Genetics)* 2:743-755) or a GAL4-VP16 fusion protein system (Wang *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96:8483-8488) is used.

In other embodiments, conditional expression of a transgene is regulated by using a recombinase system that is used to turn on or off the gene's expression by recombination in the appropriate region of the genome in which the potential drug target gene is inserted. The transgene is flanked by recombinase sites, *e.g.*, FRT sites. Such a recombinase system can be used to turn on or off expression a transgene (for review of temporal genetic switches and "tissue scissors" using recombinases, see Hennighausen & Furth, 1999, *Nature Biotechnol.* 17: 1062-63). Exclusive recombination in a selected cell type may be mediated by use of a site-specific recombinase such as Cre, FLP-wild type (wt), FLP-L or FLPe. Recombination may be effected by any art-known method, *e.g.*, the method of Doetschman *et al.* (1987, *Nature* 330: 576-78; incorporated herein by reference in its entirety); the method of Thomas *et al.*, (1986, *Cell* 44: 419-28; incorporated herein by reference in its entirety); the Cre-loxP recombination system (Sternberg and Hamilton, 1981, *J. Mol. Biol.* 150: 467-86; Lakso *et al.*, 1992, *Proc. Natl. Acad. Sci. USA* 89: 6232-36; which are both incorporated herein by reference in their entireties); the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.*, 1991, *Science* 251: 1351-55); the Cre-loxP-tetracycline control switch (Gossen and Bujard, 1992, *Proc. Natl. Acad. Sci. USA* 89: 5547-51, incorporated herein by reference in its entirety); and ligand-regulated recombinase system (Kellendonk *et al.*, 1999, *J. Mol. Biol.* 285: 175-82; incorporated herein by reference in its entirety). Preferably, the recombinase is highly active, *e.g.*, the Cre-loxP or the FLPe system, and has enhanced thermostability (Rodríguez *et al.*, 2000, *Nature Genetics* 25: 139-40; incorporated herein by reference in its entirety).

In a specific embodiment, the ligand-regulated recombinase system of Kellendonk *et al.* (1999, *J. Mol. Biol.* 285: 175-82; incorporated herein by reference in its entirety) can be used. In this system, the ligand-binding domain (LBD) of a receptor, *e.g.*, the progesterone or estrogen receptor, is fused to the Cre recombinase to increase specificity of the recombinase.

In the case of an avian, a heterologous polypeptide or polypeptides encoded by the transgenic nucleic acid may be secreted into the oviduct lumen of the mature animal and

deposited as a constituent component of the egg white into eggs laid by the animal. It is also contemplated to be within the scope of the present invention for the heterologous polypeptides to be produced in the serum of a transgenic avian.

A leaky promoter such as the CMV promoter may be operably linked to a transgene, resulting in expression of the transgene in all tissues of the transgenic avian, resulting in production of, for example, immunoglobulin polypeptides in the serum. Alternatively, the transgene may be operably linked to an avian promoter that may express the transgene in a restricted range of tissues such as, for example, oviduct cells and macrophages so that the heterologous protein may be identified in the egg white or the serum of a transgenic avian. Transgenic avians produced by the sperm-mediated transfection methods of the present invention will have the ability to lay eggs that contain one or more desired heterologous protein(s) or variant thereof.

One embodiment of the present invention, therefore, is a transgenic avian produced by the sperm-mediated transfection methods of the present invention and having a heterologous polynucleotide sequence comprising a nucleic acid insert encoding a heterologous polypeptide and operably linked to an avian lysozyme gene expression control region, the gene expression control region comprising at least one 5' matrix attachment region, an intrinsically curved DNA region, at least one transcription enhancer, a negative regulatory element, at least one hormone responsive element, at least one avian CR1 repeat element, and a proximal lysozyme promoter and signal peptide-encoding region.

Another embodiment of the present invention provides a transgenic avian further comprising a transgene with a lysozyme 3' domain.

Accordingly, the invention provides transgenic avians produced by methods of the invention as described *infra*. In preferred embodiments, the transgenic avian contains a transgene comprising a heterologous peptide coding sequence operably linked to a promoter and, in certain embodiments, other regulatory elements. In more preferred embodiments, the transgenic avians of the invention produce heterologous proteins, preferably in a tissue specific manner, more preferably such that they are deposited in the serum and, most preferably, such that the heterologous protein is deposited into the egg, particularly in the egg white. In preferred embodiments, the transgenic avians produce eggs containing greater than 5 μg, 10 μg, 50 μg, 100 μg, 250 μg, 500 μg, or 750 μg, more preferably greater than 1 mg, 2 mg, 5 mg, 10 mg, 20 mg, 50 mg, 100 mg, 200 mg, 500 mg, 700 mg, 1 gram, 2 grams, 3 grams, 4 grams or 5 grams of the heterologous protein. In preferred embodiments, the transgenic avians produce an immunoglobulin molecule and deposit the immunoglobulin in the egg or serum of the avian, and preferably, the immunoglobulin isolated from the egg or

serum specifically binds its cognate antigen. The antibody so produced may bind the antigen with the same, greater or lesser affinity than the antibody produced in a mammalian cell, such as a myeloma or CHO cell.

In specific embodiments, the transgenic avians of the invention were not produced or are not progeny of a transgenic ancestor produced using a eukaryotic viral vector, more particularly, not a retroviral vector (although, in certain embodiments, the vector may contain sequences derived from a eukaryotic viral vector, such as promoters, origins of replication, etc.). The transgenic avians of the invention include G0 avians, founder transgenic avians, G1 transgenic avians, avians containing the transgene in the sperm or ova, avians mosaic for the transgene and avians containing copies of the transgene in most or all of the cells. Contemplated by the invention are transgenic avians in which the transgene is episomal. In more preferred embodiments, the transgenic avians have the transgene integrated into one or more chromosomes. Chromosomal integration can be detected using a variety of methods well known in the art, such as, but not limited to, Southern blotting, PCR, etc.

6. EXAMPLES

The present invention is further illustrated by the following examples. Each example is provided by way of explanation of the invention, and is not intended to be a limitation of the invention. In fact, it will be apparent to those skilled in the art that various modifications, combination, additions, deletions and variations can be made in the present invention without departing from the scope or spirit of the invention. For instance, features illustrated or described as part of one embodiment can be used in another embodiment to yield a still further embodiment. It is intended that the present invention covers such modifications, combinations, additions, deletions and variations as come within the scope of the appended claims and their equivalents.

All references cited herein are incorporated herein by reference in their entirety and for all purposes to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety for all purposes. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

6.1 Example 1: Vectors Having Sperm-Specific Reporter Genes

The specific activity of spermatogenesis-specific promoters, such as the protamine promoter necessary for post-meiotic-specific transcription of this gene may be used to selectively mark those sperm cells that have inherited the transgene of interest after meiotic segregation.

The construct contains two separate elements. In one example, the first element comprises an oviduct-specific promoter, such as that associated with a gene encoding ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin or ovomucin, The promoter is operatively linked to, and therefore drives the expression of a gene coding for a desired heterologous protein of interest, such as, but not limited to, a therapeutic protein like interferon, erythropoietin (EPO), or an immunoglobulin.

The second element, which can be located either upstream or downstream from the first element, contains the protamine promoter, or any fragment thereof that is sufficient to drive the expression of a marker gene encoding a vital and color marker, such as the Green Fluorescent Protein (GFP). Those sperm cells that incorporate the transgene into their genomic DNA are vitally labeled during the late stages of spermiogenesis by the expression of the GFP protein. Given that the construct contains both the above first and the second elements, positive sperm cells also contain the transgene of interest.

Large numbers of positive sperm cells expressing the GFP protein are isolated using Fluorescent Activated Cell Sorting (FACS). Sperm cells selected on the basis of the expression of the incorporated marker gene are then used to breed hens by artificial insemination protocols. Suitable avian insemination protocols have been described by Etches (1996) Reprod. in Poultry (CAB International, Wallingford, UK), incorporated herein by reference in its entirety. In those cases where the number of positive sperm obtained after FACS isolation is too low for the likelihood of successful artificial insemination, the females may be fertilized by the intramagnal insemination method of Engel (1991) Poult. Sci. 70:1965-1969 or Trefil (1996) Br. Poult. Sci. 37:661-664, incorporated herein by reference in their entireties. Alternatively, small numbers of positive sperm cells are isolated under a microscope using UV light and then microinjected into unfertilized eggs via the Intracytoplasmic Sperm Injection (ICSI) protocols of Perry (1999), incorporated herein by reference in its entirety.

6.2 Example 2: Lipofection Gene Transfer to Avian Oocytes

(a) Isolation of the ovum: Donor hens were inseminated using the protocol for avian artificial insemination described by Etches (1996), incorporated herein by reference in

its entirety. Fertilized ova were collected from the magnum region of the oviduct of euthanized birds 1.5-3 hours after oviposition. Alternatively, a hen whose oviduct is fistulated allows the collection of eggs for enucleation as taught by Gilbert and Woodgush, (1963, *J. Reprod. Fertility* 5: 451-453) and Pancer *et al.*, (1989, *Br. Poult. Sci.* 30: 953-7). The thick albumen capsule surrounding the ovum was removed using spatulas and the ovum was placed in a well 48mm diameter and 23 mm in height containing Perry's salt solution (see Perry (1988), incorporated herein by reference in its entirety).

(b) Preparation of lipofection solutions: Two lipofection solutions were used. The first solution comprised 50μg/ml of LIPOFECTAMINE™ (Gibco) pre-incubated for 1 hour with the restriction endonuclease *Not* I (500 Units *Not* I per ml of lipofection solution), and designated herein as "Lipofectamine/Not I solution". The second lipofection solution was composed of 50μg/ml of LIPOFECTAMINE™ pre-incubated for 1 hour with 500μg of peGFP linearized with *Not* I per ml of lipofection solution, herein described as "Lipofectamine/peGFP solution." Lipofectin-treated eggs were then incubated for 1 hour.

(c) Gene transfer to avian oocytes by lipofection: The isolated ovum was then placed inside a glass conical chamber (Figure 1A) so that the blastodisc was located in the center of a window that opens at the narrower end of the conical chamber. A 40 mm diameter and 8 mm high glass dish was used at the bottom of the cone to close the system. Perry salt solution was added to the bottom of the dish to prevent drying of the lower half of the ovum. The Perry's salt solution overlaying the blastodisc (accessed through the window opening of the cone) was then replaced by, for example, 100 μl, of a lipofection solution described below. The eggs were incubated for 1 hour. Alternatively, egg incubation can be done by adding the lipofection solutions to the well and inverting the position of the incubation chamber (Figure 1B), or by using a cloning cylinder around the blastodisc (Figure 1C).

(d) Transfer of the lipofected egg: In a preferred embodiment, the ovum is surgically transferred into the oviduct of the recipient hen shortly after lipofection according to a described surgical procedure. (Tanaka, 1994, *supra*). The recipient hens are anesthetized by wing vein injection with pentobarbital (0.7 ml of a 68 mg/ml solution) or using gas anesthetics such as Isoflurane shortly after laying. During this window, the infundibulum is receptive to receiving a donor ovum but that has not yet ovulated. Feathers are removed from the abdominal area, the area is scrubbed with betadine and rinsed with 70% ethanol. The bird is placed in a supine position and a surgical drape is placed over the bird exposing the surgical area. An incision is made beginning at the junction of the sternal rib to the breastbone and running parallel to the breastbone. The length of the incision is

approximately 6cm. After cutting through the smooth muscle layers and the peritoneum, the infundibulum is located. The infundibulum is externalized and opened using gloved hands. The donor ovum is gently placed in the open infundibulum. Gravity facilitates the movement of the ovum through the infundibulum and into the anterior magnum. The internalized ovum is placed into the body cavity and the incision closed using interlocking stitches both for the smooth muscle layer and the skin. The recipient hen is returned to her cage and allowed to recover with free access to both feed and water. The hens resume normal activities after a post-operative recovery time of less than 45 minutes. Once transferred, the embryo develops inside the recipient hen and travels through the oviduct where it is encapsulated by natural egg white proteins and a natural eggshell. Eggs laid by the recipient hens are collected the next day, set, and incubated in a Jamesway incubator. The eggs hatch 21 days later.

6.3 Example 3: Maintenance of Plasmid Linearization in the Remi Procedure

A plasmid that is to be integrated into the genomic nucleic acid of a sperm is linearized by cleavage with a selected restriction endonuclease. The linearized nucleic acid is then dephosphorylated at the exposed 5' ends of the newly formed cohesive regions by alkaline phosphatase treatment. Suitable protocols for the alkaline phosphatase dephosphorylation of nucleic acids are disclosed, for example, by Sambrook *et al.*, (*supra*), incorporated herein by reference in its entirety.

While not wishing to be bound by any one theory, it is believed that dephosphorylated cohesive ends of the nucleic acid may hybridize to recircularize the cleaved plasmid. Dephosphorylation of the 5' termini, however, prevent a DNA ligase from covalently rejoining a 5' terminus to the adjacent 3' terminus, thereby preventing a stable circular plasmid molecule from reforming. The cohesive ends of the non-ligated circularized plasmid may dissociate within a sperm cell to give a linearized nucleic acid that may integrate into the sperm genomic DNA.

Alternatively, a circular plasmid having a heterologous nucleic acid that is to be integrated into the genomic nucleic acid of a sperm is digested with at least two different restriction endonucleases that generate a linearized plasmid having two non-cohesive ends, and wherein the desired transgenic element heterologous nucleic acid remains intact between the new termini of the cleaved plasmid. The restriction endonucleases are selected to give dissimilar cohesive ends that cannot hybridize together to recircularize the cleaved plasmid. The linearized nucleic acid is then delivered to the sperm with both of the

restriction endonucleases used to cleave the plasmid. The restriction endonucleases may be delivered to the sperm sequentially or simultaneously and combined, or sequentially delivered, with the cleaved plasmid.

It can be advantageous, depending upon the positions of the endonuclease cleavage sites within the plasmid relative to the desired transgene, to use two different endonucleases that produce hybridizable cohesive ends. In this case, the 5' termini may also be dephosphorylated with alkaline phosphatase as described above, to prevent religation and stabilization of the cleaved plasmid.

6.4 Example 4: Methods for Determing the SV40 Ori Requirement in SMT

To determine the requirement for the SV40 origin of replication in sperm-mediated transgenesis, 5 μg each of the plasmids p1083 (with the CMV promoter controlling heavy chain transcription) and p1086 (where the CMV promoter controls light chain transcription) were digested with *Dra* III which excises the SV40 origin of replication from the p1083 plasmid while retaining the SV40 origin of replication of the p1086 plasmid. For comparison, 5 μg each of the plasmids p1083 and p1086 were digested with the restriction endonuclease *Mlu* I that linearizes both plasmids while retaining the SV40 origin of replication in each of the respective plasmids.

Digested plasmids were used to transfect sperm. In a polystyrene tube, *Dra* III-digested plasmids p1086 and p1083 (5 μg of each) were added to 100 μl of OPTIMEM™ medium (Life Technologies, Gaithersburg, MD) and 10 μg of LIPOFECTAMINE™ liposome (Life Technologies, Gaithersburg, MD). In a separate tube, 100 units of *Dra* III restriction enzyme were added to 100 μl of OPTIMEM™ medium followed by 10 μg of LIPOFECTAMINE™. The tubes were incubated at room temperature for 30 minutes, then added to freshly collected semen containing 10^9 chicken sperm (approximately 300 μl of semen). The sperm, DNA-liposome, and restriction enzyme-liposome mixture was incubated at room temperature for 30 minutes.

Two White Leghorn hens were then artificially inseminated with 250 μl each of the transfection mixture. Eggs were collected for 7 days starting on the second day after fertilization, and set for hatch. Two weeks after hatch, serum samples were collected and assayed for human monoclonal antibodies by ELISA. The results are shown in Figure 10, wherein wing band number 3932 is the control.

6.5 Example 5: Gamma Irradiation of Chicken Sperm

Exogenous, linearized DNA can be integrated into the genome of a recipient sperm cell by cleaving the double-stranded genomic DNA by gamma irradiation of the sperm prior to lipofection thereof with the transgenic nucleic acid.

Wooster *et al.* found that rooster sperm irradiated with 12 Grays (Gy) of gamma irradiation resulted in about 43% residual fertility. (1977, *Can. J. Genet. Cytol.* 19, 437-446). Therefore, rooster semen will be irradiated with the following doses of gamma radiation: 0, 1, 5, 10, 15, and 20 Gy. A liposomal complex will consist of 10 μg of linearized DNA containing a promoter (*e.g.*, CMV, ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin, etc.) and transgene (*e.g.*, IFN, erythropoeitin, human monoclonal antibody immunoglobulin heavy and light chains, and GM-CSF, etc.) and 10 μg of LIPOFECTAMINE™ (Life Technologies, Gaithersburg, MD) will then be transfected into the irradiated sperm. After one hour, the irradiated and transfected sperm will be introduced into the hen by traditional artificial insemination procedures. Resulting laid eggs will be set and hatched, and transgene integration will be confirmed by Southern analysis of blood DNA.

6.6 Examples 6: Ovum Transfer to a Laying Hen

At the time of laying, recipient hens are anesthetized by wing vein injection with pentobarbital (0.7 ml of a 68 mg/ml solution) or by a gaseous anesthetic such as Isoflurane. Pentobarbital is the preferred anesthetic. At this time, the infundibulum is receptive to receiving a donor ovum but has not yet ovulated. Feathers are removed from the abdominal area, and the area is scrubbed with betadine, and rinsed with 70% ethanol. The bird is placed in a supine position and a surgical drape is placed over the bird with the surgical area exposed. An incision is made beginning at the junction of the sternal rib to the breastbone and running parallel to the breastbone. The length of the incision is approximately two inches. After cutting through the smooth muscle layers and the peritoneum, the infundibulum is located. The infundibulum is externalized and opened using gloved hands and the donor ovum is gently applied to the open infundibulum. The ovum is allowed to move into the infundibulum and into the anterior magnum by gravity feed. The internalized ovum is placed into the body cavity and the incision closed using interlocking stitches both for the smooth muscle layer and the skin. The recipient hen is returned to her cage and allowed to recover with free access to both feed and water. Recovery time for the bird to be up, moving and feeding is usually within 45 minutes of the operation's end. Eggs laid by

the recipient hens are collected the next day, set, and incubated. They will hatch 21 days later.

6.7 Example 7: Generation of Transgenic Chickens by Sperm-Mediated Transfection of Heterologous Nucleic Acid

Plasmid pRC/CMV-EGFP, 10 μg, was added to 100 μl of OPTIMEM™ medium (Life Technologies, Gaithersburg, MD) and 10 μg of LIPOFECTAMINE™ (Life Technologies, Gaithersburg, MD) liposomes, in a polystyrene tube. In a separate tube, 100 units of *Dra* III restriction enzyme was added to 100 μg of OPTIMEM™ medium followed by 10 μg of LIPOFECTAMINE™. As negative controls, plasmids p1086 and p1083 were used for pRC/CMV-EGFP in the transfection mixture. Tubes were incubated at room temperature for 30 minutes, then added to 10^9 freshly collected chicken sperm (approximately 300 μl of sperm). The sperm, DNA-liposome, and restriction enzyme-liposome mixture was incubated at room temperature for 30 minutes.

Two White Leghorn hens were inseminated with the transfection mixture, each hen receiving approximately 250 μl of the transfection mixture. Eggs were collected for 7 days starting on the second day after fertilization, and set for hatch.

Four days after hatching, blood drops from chicks were collected from leg veins with heparinized capillary tubes and placed on microscope slides. Blood smears were viewed with FITC illumination with an inverted microscope (Olympus IX70, 100 watt mercury lamp, HQ-FITC Band Pass Emission filter cube, excitation 480/40 nm, emission 535/50 nm, and 20X phase contrast objective). Auto-fluorescence was assessed using a TRITC filter (Olympus Modular B-MAX Filter cube, excitation 535/50 nm, emission 610/75 nm).

Two chicks that resulted from sperm transfected with pRC/CMV-EGFP had white blood cells showing green fluorescence. No fluorescence was seen when viewed with the TRITC filter, indicating that the green fluorescence was not due to auto-fluorescence. None of the control chicks, derived from sperm transferred with control plasmids, had green fluorescence in their blood.

6.8 Example 8: Sperm-Mediated Transfection of Japanese Quail Ova

Prophetic Example

Japanese Quail hens will be artificially inseminated with sperm transfected with vectors capable of expressing α-IFN, erythropoietin or a monoclonal antibody. ELISAs will be used to detect and measure the amount of an expressed transgene product in the animal's

serum and egg. As little as 15pg of α-interferon (α-IFN) or erythropoietin can be detected by this procedure.

To prepare the quail flock for artificial insemination, the females will be separated from the males. Once the isolated females are no longer laying fertile eggs and the males are consistently producing sufficient semen, the birds will be used for artificial insemination (A.I) procedures.

Sperm mediated transgenesis (SMT) of the quail will be performed with two plasmid vectors, pRC/CMV-IFNMM-SV40 and pRC/CMV-EPOMM-SV40. Transgenesis resulting in the integration of, and expression from, a heterologous nucleic acid encoding α-IFN has been used successfully in chickens with both viral-based and sperm-mediated transfer (SMT)-based systems. The second vector will carry the gene encoding for erythropoietin. This protein requires more extensive post-translational modification, i.e. four glycosylations, than does α-IFN. Both of the plasmid vectors will produce their respective expressed polypeptides in serum and *in ovo*. Assaying for α-IFN or EPO production in serum will begin at two weeks of age and egg production will occur shortly thereafter. SMT will be performed with vectors having immunoglobulin heavy and light chain under the expression control of a lysozyme promoter.

About 50 chicks will be obtained from the SMT-A.I.'s. Based on results from our chicken SMT experiments, at least 2 to 4 transgenic quail for every 50 birds will be produced from the SMT-A.I.'s.

6.9 Example 9: Preparation of Female and Male Japanese Quails for Sperm-Mediated Transfection by Artificial Insemination

The birds used will be selected for their optimal age for fertility, according to the average life history of the quail as shown in Table 1.

Table 1: Japanese Quail life history

Hatching	16-17 days
Sexual maturity	
Females	
Under current conditions	48 days
Under optimal conditions	35-38 days
Males	35-42 days
Optimal Fertility	
Females	60-240 day (8-34 weeks)
Males	60-280 days (8-40 weeks)

Declining Fertility Fertility declines 30-50%	224 days (32 weeks)
Gamete production cessation	
Females (eggs)	1.5-2 years
Males Decreases	1.5 years
	3 years
Ceases	
Lifespan	
Females	2.5-3 years
Males	3-5 years

Females: Female Japanese Quail are separated from males and eggs inspected for fertilization over a 10 day period. Females will begin producing infertile eggs about 7 to 10 days after removal from the males.

Males: Males birds will be separated from females and conditioned for semen collection. The conditioning will be continued for 10 days. About 60% of males at their sexual peak will produce good semen, and consistently high volume semen producing birds will be progressively selected.

(a) Semen Quality and Lipofection Optimization: The extracted quail sperm will be added to a diluent that is at a higher pH than is typically used with chicken sperm. Quail sperm, compared to chicken sperm, require a higher pH to maintain motility once collected from the animal, as reported by Holm. L. & Wishart G.J. in Animal Reprod. Sci. 54: 45-54 (1998) and incorporated herein by reference in its entirety. A semen diluent having a pH of between about 8 and about 9 maintains motility better than does a pH of 7.

Artificial insemination (A.I): Each hen will be artificial inseminated with a 25µl dose containing 2.5×10^7 sperm per hen. Hens will be divided into Group 1: 4 females inseminated with semen only; Group 2: 4 females inseminated with semen treated with LIPOFECTAMINE™; Group 3: 4 females sperm-mediated transfected with pCMV-IFN-SV40; Group 4: 4 females sperm-mediated transfected with pCMV-EPO-SV40. Since the average fertility of hens after artificial insemination is about 4 days, the hens will be inseminated twice a week to ensure delivery of a fresh supply of transfected semen. Transgenic positive birds will be mated to produce G_1 chicks. The first three eggs of each bird will be screened for IFN, EPO or immunoglobulin polypeptide expression and the remaining eggs will be incubated to hatching.

(b) Hatchling care: Newly hatched chicks will be grown at 105° F-110°F for the first 4-5 days. The temperature will then be reduced by 5°F after the first and second week

in 2° F increments. By the third week the house temperature will be sufficient. A 16/8 lighting schedule will also be used.

6.10 Example 10: Quail Semen Collection

The male bird is grasped so its breastbone rests in the palm of the right hand. The tail is positioned so the first two fingers of right hand lay on either side of the vent just below the legs. Holding the male in an almost vertical position, the left hand gently squeezes four times at the base of the cloaca to remove the foamy secretions of the *glandula proctodealis* (foam gland). The vent is wiped to remove traces of the foamy substance and to prevent contamination of the semen. The left hand maintains firm pressure against the base of the cloaca and gently pulls back on dorsal proctodeal wall to achieve erection.

The first two fingers of the right hand gently massage the abdomen and apply moderate pressure just below the vent to force semen from the *vas deferens* into the copulatory organ. The semen will appear shortly thereafter. The viscous, pale yellow to white semen is collected with a 20 µl pipette and immediately diluted with 150 mM NaCl and 20mM N-tris[Hydroxymethyl]methyl-2-aminoethane-sulfonic acid (TES), at pH 8.0.

6.11 Example 11: Lipofection of Quail Sperm

Quail semen will be diluted, immediately after harvesting, to a concentration of 10^8 sperm/ml in 150 mM NaCl and 20mM N-tris[Hydroxymethyl]methyl-2-aminoethane-sulfonic acid (TES), pH 8.0 buffer. Semen extender that is optimized for chicken sperm may not be used since it rapidly immobilizes quail semen within five minutes of contact.

The lipofection procedures used with quail sperm will be similar to those adopted for chicken lipofection, including REMI sperm mediated tranfections (SMT). With the chicken SMT procedure, artificial insemination is with approximately 6 x 10^8 sperm. Due to the limited amount of semen produced by male quail $1x10^8$ quail sperm will be used per hen. The lowest number of sperm that will still gives maximum insemination will be adopted. Typically, the DNA (1.0µg), restriction enzyme, LIPOFECTAMINE™ (1.0µg) and sperm (10^8) will be incubated together at a ratio of / respectively for 30 minutes. All reactions will be carried out in OptiMEM™ medium (Gibco-BRL, Gaithersburg, MD).

6.12 Integration of Adeno-Associated Virus (Aav) Inverted Terminal Repeats-Flanked Genes Introduced by Sperm-Mediated Transgenesis

The chromosomal integration of plasmid DNA into the genome of an avian cell will be mediated by flanking the gene of interest and sequences related to its expression, with

AAV inverted terminal repeat (ITR) sequences. A method for gene delivery and integration of heterologous nucleic acid sequences into the genomic DNA of a mammalian cell is described by Solis *et al.* in U.S. Patent Serial No. 5,843,742 incorporated herein by reference in its entirety. A nucleic acid segment will also be included with the gene of interest that will result in the expression of the AAV Rep protein within the same cell.

For example, a plasmid nucleic acid vector containing an expression cassette consisting of a CMV immediate early promoter driving the expression of human erythropoetin, will be flanked by AAV ITR sequences. This plasmid will be introduced by sperm-mediate transgenesis into targeted host cells together with a second nucleic acid vector plasmid. This second plasmid will include an expression cassette comprising the CMV immediate early promoter driving expression of the nucleic acid sequence encoding the AAV Rep 78 protein. Alternatively, a single nucleic acid vector comprising the expression cassette comprising the CMV immediate early promoter driving expression of the nucleic acid sequence encoding the AAV Rep 78 protein and the cassette expressing the gene of interest, such as erythropoetin, will be introduced together into an avian male gem cell.

6.13 Example 13: DNA Construct Modification to Improve Germline Transmission of Trangenes

Following genetic modification in vertebrates, a low percentage of offsprings derived from the founder animals are transgenic given the low number of germline cells that carry the transgene. As a result, costly and cumbersome breeding of the founder animals is required to expand the number of transgenic animals derived from the original founder animals.

A number of articles (*e.g.*, Peschon, 1989, *Ann. NY Acad Sci.* 564: 186-197; Peschon, 1987, *PNAS* 84: 5316-5319; Zambrowicz, 1993, *PNAS* 90: 5071; Braun, 1989, *Gene Dev.* 3:793-802; Rhim, 1995, *Biol. Reprod.* 52:20-32) as well as patent application(s) (O'Gorman *et al.*, PCT Publication No. WO 99/10488) have identified and used the elements of the protamine promoter necessary for post-meiotic-specific transcription of this gene. Other spermiogenesis-specific promoters have also been described and used in the context of genetic manipulation (Sage, 1999, *Mech. Dev.* 80: 29-39; Vidal, 1998, *Mol. Reprod. Dev.* 51: 274-280). In this example, we take advantage of the specific activity of these promoters to selectively mark those sperm cells that have inherited the transgene of interest after meiotic segregation.

In the example described here, the construct would contain two independent elements. In a preferred example, the first element would comprise an oviduct-specific promoter, such as ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin. The promoter would drive expression of a gene coding for a protein of interest, such as a therapeutic protein like Interferon, erythroprotin (EPO). Alternatively, constitutive promoters such as CMV or RSV may also be used.

The second element, located up or downstream from the first, would contain the protamine promoter, or a segment of this promoter that is sufficient to drive the expression of a marker gene. In a preferred example, the protamine promoter would drive the expression of a marker, preferably a vital and color marker, such as the Green Fluorescent Protein (GFP). In such example, those sperm cells that have inherited the transgene would be vitally labeled during the late stages of spermiogenesis with the expression of the GFP protein. Given that the construct used contains both the first and the second elements described above, positive sperm cells would also contain the transgene of interest.

Large numbers of positive sperm cells expressing the GFP proteins could be isolated using Fluorescent Activated Cell Sorting (FACS). These sperm cells could subsequently be used to breed hens by described artificial insemination protocols. (Etches, 1996, *Mol. Reprod. Dev.* 45:2918). In cases where the number of positive sperm after FACS isolation is low and insufficient for AI, the females could be bred through intramagnal insemination. (Engel, 1991, *Poultry Sci.* 70: 1965; Trefil, 1996, *Br. Poult. Sci.* 37: 661-664). Alternatively, small numbers of positive sperm cells could be isolated under a microscope using UV light and injected into unfertilized eggs via described Intracytoplasmic Sperm Injection (ICSI) protocols. (Perry, 1999, *Science* 284: 1180-83).

6.14 Example 14: Use of Chicken Centromeric and Telomeric Sequences to Create a Chicken Artificial Chromosome (ChAC)

The Shemesh *et al.* procedure (2000, *Molecular Reproduction and Development* 56: 306-308) for introducing linearized plasmid DNA into chicken sperm appears to rely on vector sequences which include an SV40 origin of replication. It is possible that the exogenous DNA therefore replicates as an episome and would most likely be lost in subsequent cell divisions due to improper segregation at mitosis. To insure proper segregation at mitosis, chicken centromere and telomere sequences could be included in the transgenic construct. Chicken centromere and telomere sequences could be obtained on a BAC (bacterial artificial chromosome) library clone from Texas A&M University or Martin Groenen at Wageningen Agricultural University, The Netherlands. The SV40 origin of

replication and the promoter (*i.e.* CMV, ovalbumin, lysozyme, ovomucoid, ovotransferrin, conalbumin, and ovomucin, etc.) and transgene (*i.e.* IFN, EPO, human monoclonal antibody heavy and light chains, GM-CSF, etc.) combination could be cloned into the BAC clone containing the chicken centromere. This BAC would therefore contain an origin of replication, a centromere, telomere, and the promoter/transgene combination which could be transfected into sperm with the Shemesh procedure. Due to the chicken centromere and telomeres, the construct would replicate and segregate as a chicken artificial chromosome (ChAC).

6.15 Example 15: Construction of Lysozyme Promoter Plasmids

The chicken lysozyme gene expression control region was isolated by PCR amplification. Ligation and reamplification of the fragments thereby obtained yielded a contiguous nucleic acid construct comprising the chicken lysozyme gene expression control region operably linked to a nucleic acid sequence optimized for codon usage in the chicken (SEQ ID NO: 5) and encoding a human interferon α2b polypeptide optimized for expression in an avian cell.

White Leghorn Chicken (*Gallus gallus*) genomic DNA was PCR amplified using the primers 5pLMAR2 (SEQ ID NO: 1) and LE-6.1kbrev1 (SEQ ID NO: 2) in a first reaction, and Lys-6.1 (SEQ ID NO: 3) and LysE1rev (SEQ ID NO: 4) as primers in a second reaction. PCR cycling steps were: denaturation at 94°C for 1 minute; annealing at 60°C for 1 minute; extension at 72°C for 6 minutes, for 30 cycles using TAQ PLUS PRECISION DNA polymerase (STRATAGENE®, LaJolla, CA). The PCR products from these two reactions were gel purified, and then united in a third PCR reaction using only 5pLMAR2 (SEQ ID NO: 1) and LysE1rev (SEQ ID NO: 4) as primers and a 10-minute extension period. The resulting DNA product was phosphorylated, gel-purified, and cloned into the *Eco*R V restriction site of the vector pBLUESCRIPT® KS, resulting in the plasmid p12.0-lys.

p12.0-lys was used as a template in a PCR reaction with primers 5pLMAR2 (SEQ ID NO: 1) and LYSBSU (5'-CCCCCCCCTAAGGCAGCCAGGGGCAGGAAGCAAA-3') (SEQ ID NO: 5) and a 10 minute extension time. The resulting DNA was phosphorylated, gel-purified, and cloned into the *Eco*R V restriction site of pBLUESCRIPT® KS, forming plasmid p12.0lys-B.

p12.0lys-B was restriction digested with *Not* I and *Bsu*36 I, gel-purified, and cloned into *Not* I and *Bsu*36 I digested pCMV-LysSPIFNMM, resulting in p12.0-lys-LSPIFNMM. p12.0-lys-LSPIFNMM was digested with *Sal* I and the SalItoNotI primer (5'-TCGAGCGGCCGC-3') (SEQ ID NO: 13) was annealed to the digested plasmid, followed

by *Not* I digestion. The resulting 12.5 kb *Not* I fragment, comprising the lysozyme promoter region linked to IFNMAGMAX-encoding region and an SV40 polyadenylation signal sequence, was gel-purified and ligated to *Not* I cleaved and dephosphorylated PBLUESCRIPT® KS, thereby forming the plasmid pAVIJCR-A115.93.1.2, which was then sequenced.

6.16 Example 16: Construction of Plasmids Which Contain the 3' Lysozyme Domain

The plasmid pAVIJCR-A115.93.1.2 (containing the -12.0 kb lysozyme promoter controlling expression of human interferon α2b) was purified with a QIAGEN® Plasmid Maxi Kit (QIAGEN®, Valencia, CA), and 100 µg of the plasmid were restriction digested with *Not*I restriction enzyme. The digested DNA was phenol/$CHCl_3$ extracted and ethanol precipitated. Recovered DNA was resuspended in 1mM Tris-HCl (pH 8.0) and 0.1mM EDTA, then placed overnight at 4°C. DNA was quantified by spectrophotometry and diluted to the appropriate concentration. The DNA samples were bound to the SV40 T antigen NLS peptide by incubation for 15 minutes.

The plasmid pAVIJCR-A115.93.1.2 was restriction digested with *Fse*I and blunt-ended with T4 DNA polymerase. The linearized, blunt-ended pAVIJCR-A115.93.1.2 plasmid was then digested with *Xho*I restriction enzyme, followed by treatment with alkaline phosphatase. The resulting 15.4 kb DNA band containing the lysozyme 5' matrix attachment region (MAR) and –12.0 kb lysozyme promoter driving expression of a human interferon was gel purified by electroelution.

The plasmid pIIIilys was restriction digested with *Mlu*I, then blunt-ended with the Klenow fragment of DNA polymerase. The linearized, blunt-ended pIIIilys plasmid was digested with *Xho*I restriction enzyme and the resulting 6 kb band containing the 3' lysozyme domain from exon 3 to the 3' end of the 3' MAR was gel purified by electroelution. The 15.4 kb band from pAVIJCR-A115.93.1.2 and the 6 kb band from pIIIilys were ligated with T4 DNA ligase and transformed into STBL4 cells (Invitrogen Life Technologies, Carlsbad, CA) by electroporation. The resulting 21.3 kb plasmids from two different bacterial colonies were named pAVIJCR-A212.89.2.1 and pAVIJCR-A212.89.2.3 respectively.

6.17 Example 17: Construction of an ALV-based Vector Having β-lactamase Encoding Sequences

The *lacZ* gene of pNLB, a replication-deficient avian leukosis virus (ALV)-based vector (Cosset *et al.*, *J. Virol.* 65: 3388-94 (1991)), was replaced with an expression cassette consisting of a cytomegalovirus (CMV) promoter and the reporter gene *β-lactamase* (*β-La* or *BL*).

To efficiently replace the *lacZ* gene of pNLB with a transgene, an intermediate adaptor plasmid was first created, pNLB-Adapter. pNLB-Adapter was created by inserting the chewed back *ApaI*/*ApaI* fragment of pNLB (Cosset *et al.*, 1991, *J. Virol.* 65:3388-94) (in pNLB, the 5′ *ApaI* sites reside 289 bp upstream of *lacZ* and the 3′ *ApaI* sites reside 3′ of the 3′ LTR and Gag segments) into the chewed-back *KpnI*/*SacI* sites of pBLUESCRIPT®KS(-). The filled-in *MluI*/*XbaI* fragment of pCMV-BL (Moore *et al.*, *Anal. Biochem.* 247: 203-9 (1997)) was inserted into the chewed-back *KpnI*/*NdeI* sites of pNLB-Adapter, replacing *lacZ* with the CMV promoter and the *BL* gene (in pNLB, *KpnI* resides 67 bp upstream of *lacZ* and *NdeI* resides 100 bp upstream of the *lacZ* stop codon), thereby creating pNLB-Adapter-CMV-BL. To create pNLB-CMV-BL, the *HindIII*/*BlpI* insert of pNLB (containing *lacZ*) was replaced with the *HindIII*/*BlpI* insert of pNLB-Adapter-CMV-BL. This two step cloning was necessary because direct ligation of blunt-ended fragments into the *HindIII*/*BlpI* sites of pNLB yielded mostly rearranged subclones, for unknown reasons.

6.18 Example 18: Production of Transduction Particles Having an ALV-based Vector Having β-lactamase Encoding Sequences

Sentas and Isoldes were cultured in F10 (GIBCO®), 5% newborn calf serum (GIBCO®), 1% chicken serum (GIBCO®), 50 μg/ml phleomycin (Cayla Laboratories) and 50 μg/ml hygromycin (SIGMA®). Transduction particles were produced as described in Cosset *et al.*, 1991, herein incorporated by reference, with the following exceptions. Two days after transfection of the retroviral vector pNLB-CMV-BL (from Example 10, above) into 9 x 10^5 Sentas, virus was harvested in fresh media for 6-16 hours and filtered. All of the media was used to transduce 3 x 10^6 Isoldes in three 100 mm plates with polybrene added to a final concentration of 4 μg/ml. The following day the media was replaced with media containing 50 μg/ml phleomycin, 50 μg/ml hygromycin and 200 μg/ml G418 (SIGMA®). After 10-12 days, single $G418^r$ colonies were isolated and transferred to 24-well plates. After 7-10 days, titers from each colony was determined by transduction of Sentas followed by G418 selection. Typically 2 out of 60 colonies gave titers at 1-3 x 10^5. Those colonies were expanded and the virus concentrated to 2-7 x 10^7 as described in

Allioli *et al.*, 1994, *Dev. Biol.* 165:30-7, herein incorporated by reference. The integrity of the CMV-BL expression cassette was confirmed by assaying for β-lactamase in the media of cells transduced with NLB-CMV-BL transduction particles.

6.19 Example 19: pNLB-CMV-IFN Vector Having an IFN Encoding Sequence

The DNA sequence for human interferon α2b based on hen oviduct optimized codon usage was created using the BACKTRANSLATE program of the Wisconsin Package, version 9.1 (Genetics Computer Group. Inc., Madison, WI) with a codon usage table compiled from the chicken (*Gallus gallus*) ovalbumin, lysozyme, ovomucoid, and ovotransferrin proteins. The template and primer oligonucleotides (SEQ ID NOS: 14-31) shown in Figures 8A-B were amplified by PCR with *Pfu* polymerase (STRATAGENE®, La Jolla, CA) using 20 cycles of 94°C for 1 min., 50°C for 30 sec., and 72°C for 1 min. and 10 sec.

PCR products were purified from a 12% polyacrylamide-TBE gel by the "crush and soak" method (Maniatis *et al.* 1982), then combined as templates in an amplification reaction using only IFN-1 (SEQ ID NO: 21) and IFN-8 (SEQ ID NO: 31) as primers. The resulting PCR product was digested with *Hind* III and *Xba* I and gel purified from a 2% agarose-TAE gel, then ligated into *Hind* III and *Xba* I digested, alkaline phosphatase-treated, pBLUESCRIPT® KS (STRATAGENE®), resulting in the plasmid pBluKSP-IFNMagMax. Both strands were sequenced by cycle sequencing on an ABI PRISM 377 DNA Sequencer (Perkin-Elmer, Foster City, CA) using universal T7 or T3 primers. Mutations in pBluKSP-IFN derived from the original oligonucleotide templates were corrected by site-directed mutagenesis with the Transformer Site-Directed Mutagenesis Kit (Clontech, Palo Alto, CA). The interferon coding sequence was then removed from the corrected pBluKSP-IFN with *Hind* III and *Xba* 1, purified from a 0.8% agarose-TAE Gel, and ligated to *Hind* III and *Xba* I digested, alkaline phosphatase-treated pCMV-BetaLa-3B-dH. The resulting plasmid was pCMV-IFN which contained IFN coding sequence controlled by the cytomegalovirus immediate early promoter/enhancer and SV40 polyA site.

To clone the IFN coding sequence controlled by the CMV promoter/enhancer into the NLB retroviral plasmid, pCMV-IFN was first digested with *Cla*I and *Xba*I, then both ends were filled in with Klenow fragment of DNA polymerase (New England BioLabs, Beverly, MA). pNLB-adapter was digested with *Nde* I and *Kpn* I, and both ends were made blunt by T4 DNA polymerase (New England BioLabs). Appropriate DNA fragments were

purified on a 0.8% agarose-TAE gel, then ligated and transformed into DH5α cells. The resulting plasmid was pNLB-adapter-CMV-IFN.

This plasmid was then digested with *Mlu* I and partially digested with *Blp* I and the appropriate fragment was gel purified. pNLB-CMV-EGFP was digested with *Mlu* I and *Blp* I, then alkaline-phosphatase treated and gel purified. The *Mlu* I/*Blp* I partial fragment of pNLB-adapter-CMV-IFN was ligated to the large fragment derived from the *Mlu* I/*Blp* I digest of pNLB-CMV-EGFP, creating pNLB-CMV-IFN.

6.20 Example 20: Production of pNLB-CMV-IFN Transduction Particles

Senta packaging cells (Cosset *et al.*, 1991) were plated at a density of 3×10^5 cells/35 mm tissue culture dish in F-10 medium (Life Technologies) supplemented with 50% calf serum (Atlanta Biologicals), 1% chicken serum (Life Technologies), 50 μg/ml hygromycin (SIGMA®), and 50 μg/ml phleomycin (CAYLA, Toulouse, France). These cells were transfected 24h after plating with 2 μg of CsCl-purified pNLB-CMV-IFN DNA and 6 μl of Lipofectin liposomes (Life Technologies) in a final volume of 500 μl Optimem (Life Technologies). The plates were gently rocked for four hours at 37° C in a 5% CO_2 incubator. For each well, the media was removed, washed once with 1 ml of Optimem and re-fed with 2 mls of F-10 medium supplemented with 50% calf serum, 1% chicken serum, 50 μg/ml hygromycin, and 50 μg/ml phleomycin. The next day, medium from transfected Sentas was recovered and filtered through a 0.45 micron filter.

This medium was then used to transduce Isolde cells. 0.3 ml of the filtered medium recovered from Senta cells was added to 9.6 ml of F-10 (Life Technologies) supplemented as described above, in addition to polybrene (SIGMA®) at a final concentration of 4 μg/ml. This mixture was added to 10^6 Isolde packaging cells (Cosset *et al.*, 1991) plated on a 100mm dish the previous day, then replaced with fresh F-10 medium (as described for Senta growth) 4 hours later.

The next day, the medium was replaced with fresh medium which also contained 200 μg/ml neomycin (G418, SIGMA®). Every other day, the medium was replaced with fresh F-10 medium supplemented with 50% calf serum, 1% chicken serum, 50 μg/ml hygromycin, 50 μg/ml phleomycin, and 200 μg/ml neomycin. Eleven to twelve days later, single colonies were visible by eye, and these were picked and placed into 24 well dishes. When some of the 24 well dishes became confluent, medium was harvested and titered to determine the cell lines with the highest production of retrovirus.

Titering was performed by plating 7.5×10^4 Senta cells per well in 24 well plates on the day prior to viral harvest and transduction. The next day 1 ml of fresh F-10 medium

supplemented with 50% calf serum, 1% chicken serum, 50 μg/ml hygromycin, and 50 μg/ml phleomycin was added to each well of the isolated Isolde colonies. Virus was harvested for 8-10 hours. The relative density of each well of Isoldes was noted. After 8-10 hours, 2 and 20μl of media from each well of Isoldes was added directly to the media of duplicate wills of the Sentas. Harvested medium was also tested for the presence of interferon by IFN ELISA and for interferon bioreactivity. The next day the media was replaced with F-10 medium supplemented with 50% calf serum, 1% chicken serum, 50 μg/ml hygromycin, 50 μg/ml phleomycin, and 200 μg/ml neomycin. When obvious neomycin-resistant colonies were evident in the wells of transduced Sentas, the number of colonies was counted for each well.

The Isolde colony producing the highest titer was determined by taking into account the number of colonies and correcting for the density of the Isolde cells when the viral particles were harvested (i.e., if two Isolde colonies gave rise to media with the same titer, but one was at a 5% density and the other was at a 50% density at the time of viral harvest, the one at the 5% density was chosen for further work, as was the case in the present example).

The Isolde cell line producing the highest titer of IFN-encoding transducing particles was scaled up to six T-75 tissue culture flasks. When flasks were confluent, cells were washed with F-10 medium (unsupplemented) and transducing particles were then harvested for 16 hours in 14 ml/flask of F-10 containing 1% calf serum (Atlanta Biologicals) and 0.2% chicken serum (Life Technolocyies). Medium was harvested, filtered through a 0.45 micron syringe filter, then centrifuged at 195,000xg in a Beckman 60Ti rotor for 35 min. Liquid was removed except for 1 ml, and this was incubated with the pellet at 37°C with gentle shaking for one hour. Aliquots were frozen at -70°C. Transducing particles were then titered on Senta cells to determine concentrations used to inject avian sperms.

6.21 Example 21: Construction of Lysozyme Promoter Plasmids

The chicken lysozyme gene expression control region isolated by PCR amplification is fully disclosed in U.S. Patent Application Serial No. 09/922,549, filed August 3, 2001 and incorporated herein by reference in its entirety. Ligation and reamplification of the fragments thereby obtained yielded a functionally contiguous nucleic acid construct comprising the chicken lysozyme gene expression control region operably linked to a nucleic acid sequence encoding a human interferon α2b polypeptide and optimized for codon usage in the chicken. Briefly, chicken (*Gallus gallus* (White Leghorn)) genomic DNA was PCR amplified using the primers 5pLMAR2 and LE-6.1kbrev1 in a first

reaction, and Lys-6.1 and LysE1rev as primers in a second reaction. PCR cycling steps were: denaturation at 94°C for 1 minute; annealing at 60°C for 1 minute; extension at 72°C for 6 minutes, for 30 cycles using TAQ PLUS PRECISION™ DNA polymerase (STRATAGENE®, La Jolla, CA). The PCR products from these two reactions were gel purified, and then united in a third PCR reaction using only 5pLMAR2 and LysE1rev as primers and a 10 minute extension period. The resulting DNA product was phosphorylated, gel-purified, and cloned into the *Eco*R V restriction site of the vector pBLUESCRIPT® KS, resulting in the plasmid p12.0-lys.

p12.0-lys was used as a template in a PCR reaction with primers 5pLMAR2 and LYSBSU and a 10 minute extension time. The resulting DNA was phosphorylated, gel-purified, and cloned into the *Eco*R V restriction site of pBLUESCRIPT® KS, forming plasmid p12.0lys-B.

p12.0lys-B was restriction digested with *Not* I and *Bsu*36 I, gel-purified, and cloned into *Not* I and *Bsu*36 I digested pCMV-LysSPIFNMM, resulting in p12.0-lys-LSPIFNMM. p12.0-lys-LSPIFNMM was digested with *Sal* I and the SalItoNotI primer was annealed to the digested plasmid, followed by *Not* I digestion. The resulting 12.5 kb *Not* I fragment, comprising the lysozyme promoter region linked to IFNMAGMAX-encoding region and an SV40 polyadenylation signal sequence, was gel-purified and ligated to *Not* I cleaved and dephosphorylated pBLUESCRIPT® KS, thereby forming the plasmid pAVIJCR-A115.93.1.2.

6.22 Example 22: Complete Lysozyme Promoter and IFNMAGMAX Sequences

The complete sequences of the lysozyme gene promoter and the codon-optimized human interferon α2b nucleic acid are fully disclosed in U.S. Patent Application No. 09/922,549, filed 03 August 2001 and incorporated herein by reference in its entirety. The complete nucleotide sequence of the approximately 12.5 kb chicken lysozyme promoter region/IFNMAGMAX construct spans the 5′ matrix attachment region (5′ MAR), through the lysozyme signal peptide, to the sequence encoding the gene IFNMAGMAX and the subsequent polyadenylation signal sequence. The IFNMAGMAX nucleic acid sequence had been synthesized as described in Example 21 above. The expressed IFN α2b sequence within plasmid pAVIJCR-A115.93.1.2 functioned as a reporter gene for lysozyme promoter activity. This plasmid construct may also be used for production of interferon α2b in the egg white of transgenic chickens.

6.23 Example 23: Synthesis of the MDOT promoter construct

Amplification of the ovomucoid and ovotransferrin promoter sequences

Oligonucleotide primers 1 (SEQ ID NO: 32) and 2 (SEQ ID NO: 33), as shown in Figure 9 were used to amplify the ovomucoid sequences. Oligonucleotide primers 3 (SEQ ID NO: 34) and 4 (SEQ ID NO: 35) were used to amplify the ovotransferrin sequence by PCR. The primers were designed such that the PCR-amplified ovomucoid sequences contained an *Xho* I restriction cleavage site at the 5' end and a *Cla* I site at the 3′ end. Similarly, the PCR-amplified ovotransferrin product had a *Cla* I restriction site at the 5′ end and a *Hin*d III site at the 3′ end. The overlapping *Cla* I site was used to splice the two-PCR products to create the MDOT promoter construct. The nucleic acid sequence SEQ ID NO: 11 of the MDOT promoter construct is shown in Figure 11. The final product was cloned in a bluescript vector between the *Xho* I and *Hin*d III sites. From the bluescript vector the promoter region was released by *Kpn* I/*Hin*d III restriction digestion and cloned into the prc-CMV-IFN vector to replace the CMV promoter to create MDOT-IFN (clone #10). This plasmid was tested *in vitro*.

6.24 Example 24: Testicular Injection

5 weeks old White Leghorn male chickens were anesthetized using Isoflourane. Small incision was made between the last two ribs to expose the testes. A 5-10 µl virus suspension of pLNHX-CMV-EGFP/VSVg (9×10^6 per ml) was injected into either both testes or only one of the testes.

At 20 weeks of age, semen samples were collected. Only one bird had sperm in his semen. Genomic DNA was isolated from the semen and used to amplify the transgene (CMV-EGFP) by PCR reaction using different DMSO concentrations. The samples were separated on agarose gel, transferred onto nitrocellulose membrane and hybridized with EGFP probe. As shown in Figure 11, EGFP positive bands are detected at two different DMSO concentrations suggesting that (1) specific PCR conditions are required for the amplification of the transgene and (2) the sperm samples have incorporated the transgene in their genome.

EQUIVALENTS

Reference now will be made in detail to the various embodiments of the invention, one or more examples of which are illustrated in the accompanying drawings. Each example is provided by way of explanation of the invention, not limitation of the invention. In fact, it will be apparent to those skilled in the art that various modifications,

combinations, additions, deletions and variations can be made in the present invention without departing from the scope or spirit of the invention. For instance, features illustrated or described as part of one embodiment can be used in another embodiment to yield a still further embodiment. It is intended that the present invention covers such modifications, combinations, additions, deletions and variations as fall within the scope of the appended claims and their equivalents.

What is Claimed Is:

1. A method of generating a transgenic avian zygote by sperm-mediated transfection, said method comprising:

(a) obtaining a suspension of avian male germ cells selected from the group consisting of spermatozoa and spermatozoal precursor cells;

(b) introducing a nucleic acid comprising a transgene comprising a nucleotide sequence encoding a heterologous polypeptide to the avian male germ cells by lipofection, electroporation or restriction enzyme mediated integration;

(c) delivering the avian male germ cells having the nucleic acid to an avian oocyte ,

thereby generating a transgenic avian zygote having the nucleic acid incorporated therein.

2. The method of Claim 1, wherein the avian male germ cells and the avian oocyte are obtained from a chicken.

3. The method of Claim 1, wherein the avian male germ cells and the avian oocytes are obtained from a quail.

4. The method of Claim 1, wherein the nucleotide sequence encoding said heterologous polypeptide is operably linked to a transcriptional regulatory element that can direct gene expression in one or more cells of said transgenic avian.

5. The method of Claim 4, wherein the transcriptional regulatory element is selected from the group consisting of the promoter regions of the avian genes encoding ovalbumin, lysozyme, ovomucoid, ovomucin, conalbumin and ovotransferrin.

6. The method of Claim 5, wherein the selected nucleic acid further comprises a chicken lysozyme gene expression controlling region comprising the nucleotide sequence of SEQ ID NO: 7.

7. The method of Claim 4, wherein the transcriptional regulatory element is a tissue specific promoter.

8. The method of Claim 7, wherein the tissue specific promoter is specific for the magnum.

9. The method of Claim 1, wherein the transgene comprises at least one cytomegalovirus promoter.

10. The method of Claim 9, wherein the transcriptional regulatory element comprises at least two regions derived from the promoter of an avian gene, said regions being from a different promoter.

11. The method of Claim 10, wherein the transcriptional regulatory element has the nucleotide sequence of SEQ ID NO: 11.

12. The method of Claim 1, wherein the transgene comprises at least one matrix attachment region (MAR).

13. The method of Claim 12, wherein the transgene comprises a 5' MAR and a 3' MAR which flank said nucleotide sequence.

14. The method of Claim 1, wherein the heterologous polypeptide is selected from the group consisting of a cytokine, a hormone, an enzyme, a structural polypeptide, and an immuoglobulin polypeptide.

15. The method of Claim 14, wherein the cytokine is selected from the group consisting of interferon, interleukin, granulocyte colony-stimulating factor, granulocyte-macrophage colony-stimulating factor, stem cell factor, erythropoietin, thrombopoietin, and stem cell factor.

16. The method of Claim 15, wherein the cytokine is an interferon.

17. The method of Claim 1, wherein the transgene comprises an internal ribosome entry site (IRES).

18. The method of Claim 17, wherein the transgene comprises at least two nucleotide sequences each encoding a heterologous polypeptide.

19. The method of Claim 18, wherein the at least two nucleotide sequences encode at least two heterologous peptides that form a multimeric protein.

20. The method of Claim 19, wherein the multimeric protein specifically binds a selected ligand.

21. The method of Claim 20, wherein the multimeric protein is an antibody.

22. The method of Claim 1, wherein the heterologous polypeptide comprises a peptide region suitable for the isolation of the heterologous polypeptide.

23. The method of Claim 1, wherein the nucleic acid is a eukaryotic viral vector.

24. The method of Claim 23, wherein the eukaryotic viral vector is derived from any of the group consisting of avian leukosis virus, adenovirus, transferrin-polylysine enhanced adenoviral vectors, human immunodeficiency virus vectors, lentiviral vectors, and Moloney murine leukemia virus-derived vectors.

25. The method of Claim 1, wherein the nucleic acid is a plasmid vector.

26. The method of Claim 1, wherein the nucleic acid is a bacterial artificial chromosome (BAC).

27. The method of Claim 1, wherein the nucleic acid is not a eukaryotic viral vector.

28. The method of Claim 4, wherein the transcriptional regulatory element is a regulatable promoter.

29. The method of Claim 6, wherein the selected nucleic acid further comprises a region encoding the 3' region of the chicken lysozyme gene and having the nucleotide sequence of SEQ ID NO: 9.

30. The method of Claim 1, wherein the nucleotide sequence encoding said heterologous polypeptide comprises an origin of replication.

31. The method of Claim 30, wherein the origin of replication is the SV40 origin of replication.

32. The method of Claim 1, wherein the nucleic acid is selected from the group consisting of a linear nucleic acid, a plasmid, a viral nucleic acid, and an artificial chromosome.

33. The method of Claim 32, wherein the artificial chromosome further comprises a centromere and optionally a telomere.

34. The method of Claim 32, wherein the linear nucleic acid has at least one cohesive end characterized by the cohesive end generated by a restriction endonuclease.

35. The method of Claim 32, wherein the linear nucleic acid has at least one blunt end.

36. The method of Claim 34, wherein the at least one cohesive end is generated by chemical synthesis.

37. The method of Claim 34, wherein the at least one cohesive end is generated by an enzyme other than a restriction endonuclease.

38. The method of Claim 34, wherein the at least one cohesive end is generated by a combination of chemical and enzymatic methods.

39. The method of Claim 1, wherein the nucleic acid is introduced to the avian male germ cells by restriction enzyme mediated integration.

40. The method of Claim 39, further comprising the step of delivering to the avian male germ cells a restriction endonuclease capable of cleaving the genomic nucleic acid of the avian male germ cells.

41. The method of Claim 40, wherein the nucleic acid is delivered sequentially with the restriction endonuclease to the avian male germ cells.

42. The method of Claim 1, wherein the nucleic acid is delivered to the avian male germ cells by adeno-associated virus-derived vector.

43. The method of Claim 42, wherein the nucleic acid is bounded by inverted terminal repeat sequences.

44. The method of Claim 42, wherein the nucleic acid is bounded by inverted terminal repeat sequences derived from an adeno-associated virus-derived vector.

45. The method of Claim 42, wherein the adeno-associated virus-derived vector further comprises a transcription cassette capable of expressing an adeno-associated virus Rep protein.

46. The method of Claim 45, wherein the Rep protein is Rep 78.

47. The method of Claim 45, wherein the nucleic acid bounded by inverted terminal repeat sequences is inserted in a first nucleic acid vector and the transcription cassette capable of expressing an adeno-associated virus Rep protein is inserted in a second nucleic acid vector.

48. The method of Claim 1, further comprising the step of irradiating the avian male germ cells, thereby cleaving the nuclic acid, wherein the radiation is selected from the group consisting of ultraviolet light, gamma rays, X-rays, and ultrasound.

49. The method of Claim 1, wherein the avian oocyte is an isolated oocyte, and wherein the avian male germ cells having the nucleic acid are delivered to the isolated oocyte by a method selected from the group consisting of microinjection, intracytoplasmic sperm injection (ICSI), and artificial insemination.

50. The method of Claim 49, wherein the avian male germ cells having the nucleic acid therein are delivered to the nucleus of the oocyte.

51. The method of Claim 1, wherein the nucleic acid forms an episome in the avian male germ cells.

52. The method of Claim 1, wherein the nucleic acid in the avian oocyte is an episome.

53. The method of Claim 1, further comprising isolating an avian oocyte from the female of an avian by:

(a) removing an ovum from a bird after ovulation and before fertilization; and

(b) removing an albumen layer from the ovum.

54. The method of Claim 1, further comprising the steps of:

(a) fistulating an avian female;

(b) delivering the transgenic avian zygote to the infundibulum of the avian female such that said transgenic avian zygote is subsequently laid by said avian female as a shelled egg; and

(c) incubating the shelled egg until said shelled egg hatches,

thereby producing a transgenic avian containing the transgene.

55. The method of Claim 54, wherein the heterologous polypeptide is expressed in one or more cells of said transgenic avian .

56. The method of Claim 55, wherein the heterologous polypeptide is expressed in the serum of said transgenic avian.

57. The method of Claim 55, wherein the heterologous polypeptide is expressed in the magnum of said transgenic avian.

58. The method of Claim 54, further comprising the step of allowing the transgenic avian to develop to sexual maturity.

59. The method of Claim 58, wherein the heterologous polypeptide is delivered to the white of a developing avian egg produced by the transgenic avian.

60. The method of Claim 55 or 59 further comprising isolating said heterologous polypeptide from said transgenic avian or an egg produced by the transgenic avian.

61. A transgenic avian that produces at least one heterologous polypeptide in egg white, wherein the transgenic avian or founder ancestor of said transgenic avian was not produced using a eukaryotic viral vector.

62. A transgenic avian produced by the method of Claim 54.

63. The transgenic avian of Claim 61 or 62, wherein the avian is a chicken.

64. The transgenic avian of Claim 63, wherein the heterologous polypeptide is selected from the group consisting of a cytokine, a hormone, an enzyme, a structural protein, and an immunglobulin polypeptide.

65. The transgenic avian of Claim 63, wherein the cytokine is an interferon.

66. The transgenic avian of Claim 61 or 62, wherein the transgenic avian produces a heterologous multimeric protein.

67. The transgenic avian of Claim 66, wherein the heterologous multimeric protein specifically binds a selected ligand.

68. The transgenic avian of Claim 66, wherein the heterologous multimeric protein is an antibody.

69. An avian egg produced by the transgenic avian of Claim 61 or 62.

70. An avian egg produced by the transgenic avian of any of Claims 63-68.

71. A heterologous protein heterologous protein produced by the transgenic avian of Claim 61 or 62, wherein the heterologous protein comprises a heterologous polypeptide selected from the group consisting of a cytokine, a hormone, an enzyme, a structural protein, and an immunoglobulin polypeptide.

72. The heterologous polypeptide of Claim 71, wherein the cytokine is an interferon.

73. The heterologous protein of Claim 71, wherein the heterologous protein is a multimeric protein.

74. The heterologous protein of Claim 71, wherein the heterologous protein is an antibody.

SEQ ID NO: 6

TGCCGCCTTC TTTGATATTC ACTCTGTTGT ATTTCATCTC TTCTTGCCGA TGAAAGGATA 60
TAACAGTCTG TATAACAGTC TGTGAGGAAA TACTTGGTAT TTCTTCTGAT CAGTGTTTTT 120
ATAAGTAATG TTGAATATTG GATAAGGCTG TGTGTCCTTT GTCTTGGGAG ACAAAGCCCA 180
CAGCAGGTGG TGGTTGGGGT GGTGGCAGCT CAGTGACAGG AGAGGTTTTT TTGCCTGTTT 240
TTTTTTTTTT TTTTTTTTTT AAGTAAGGTG TTCTTTTTTC TTAGTAAATT TTCTACTGGA 300
CTGTATGTTT TGACAGGTCA GAAACATTTC TTCAAAAGAA GAACCTTTTG GAAACTGTAC 360
AGCCCTTTTC TTTCATTCCC TTTTTGCTTT CTGTGCCAAT GCCTTTGGTT CTGATTGCAT 420
TATGGAAAAC GTTGATCGGA ACTTGAGGTT TTTATTTATA GTGTGGCTTG AAAGCTTGGA 480
TAGCTGTTGT TACACGAGAT ACCTTATTAA GTTTAGGCCA GCTTGATGCT TTATTTTTTC 540
CCTTTGAAGT AGTGAGCGTT CTCTGGTTTT TTTCCTTTGA AACTGGTGAG GCTTAGATTT 600
TTCTAATGGG ATTTTTTACC TGATGATCTA GTTGCATACC CAAATGCTTG TAAATGTTTT 660
CCTAGTTAAC ATGTTGATAA CTTCGGATTT ACATGTTGTA TATACTTGTC ATCTGTGTTT 720
CTAGTAAAAA TATATGGCAT TTATAGAAAT ACGTAATTCC TGATTTCCTT TTTTTTTATC 780
TCTATGCTCT GTGTGTACAG GTCAAACAGA CTTCACTCCT ATTTTTATTT ATAGAATTTT 840
ATATGCAGTC TGTCGTTGGT TCTTGTGTTG TAAGGATACA GCCTTAAATT TCCTAGAGCG 900
ATGCTCAGTA AGGCGGGTTG TCACATGGGT TCAAATGTAA AACGGGCACG TTTGGCTGCT 960
GCCTTCCCGA GATCCAGGAC ACTAAACTGC TTCTGCACTG AGGTATAAAT CGCTTCAGAT 1020
CCCAGGGAAG TGCAGATCCA CGTGCATATT CTTAAAGAAG AATGAATACT TTCTAAAATA 1080
TTTTGGCATA GGAAGCAAGC TGCATGGATT TGTTTGGGAC TTAAATTATT TTGGTAACGG 1140
AGTGCATAGG TTTTAAACAC AGTTGCAGCA TGCTAACGAG TCACAGCGTT TATGCAGAAG 1200
TGATGCCTGG ATGCCTGTTG CAGCTGTTTA CGGCACTGCC TTGCAGTGAG CATTGCAGAT 1260
AGGGGTGGGG TGCTTTGTGT CGTGTTCCCA CACGCTGCCA CACAGCCACC TCCCGGAACA 1320
CATCTCACCT GCTGGGTACT TTTCAAACCA TCTTAGCAGT AGTAGATGAG TTACTATGAA 1380
ACAGAGAAGT TCCTCAGTTG GATATTCTCA TGGGATGTCT TTTTTCCCAT GTTGGGCAAA 1440
GTATGATAAA GCATCTCTAT TTGTAAATTA TGCACTTGTT AGTTCCTGAA TCCTTTCTAT 1500
AGCACCACTT ATTGCAGCAG GTGTAGGCTC TGGTGTGGCC TGTGTCTGTG CTTCAATCTT 1560
TTAAAGCTTC TTTGGAAATA CACTGACTTG ATTGAAGTCT CTTGAAGATA GTAAACAGTA 1620
CTTACCTTTG ATCCCAATGA AATCGAGCAT TTCAGTTGTA AAAGAATTCC GCCTATTCAT 1680
ACCATGTAAT GTAATTTTAC ACCCCCAGTG CTGACACTTT GGAATATATT CAAGTAATAG 1740
ACTTTGGCCT CACCCTCTTG TGTACTGTAT TTTGTAATAG AAAATATTTT AAACTGTGCA 1800
TATGATTATT ACATTATGAA AGAGACATTC TGCTGATCTT CAAATGTAAG AAAATGAGGA 1860
GTGCGTGTGC TTTTATAAAT ACAAGTGATT GCAAATTAGT GCAGGTGTCC TTAAAAAAAA 1920
AAAAAAAAAG TAATATAAAA AGGACCAGGT GTTTTACAAG TGAAATACAT TCCTATTTGG 1980
TAAACAGTTA CATTTTTATG AAGATTACCA GCGCTGCTGA CTTTCTAAAC ATAAGGCTGT 2040
ATTGTCTTCC TGTACCATTG CATTTCCTCA TTCCCAATTT GCACAAGGAT GTCTGGGTAA 2100
ACTATTCAAG AAATGGCTTT GAAATACAGC ATGGGAGCTT GTCTGAGTTG GAATGCAGAG 2160
TTGCACTGCA AAATGTCAGG AAATGGATGT CTCTCAGAAT GCCCAACTCC AAAGGATTTT 2220
ATATGTGTAT ATAGTAAGCA GTTTCCTGAT TCCAGCAGGC CAAAGAGTCT GCTGAATGTT 2280
GTGTTGCCGG AGACCTGTAT TTCTCAACAA GGTAAGATGG TATCCTAGCA ACTGCGGATT 2340
TTAATACATT TTCAGCAGAA GTACTTAGTT AATCTCTACC TTTAGGGATC GTTTCATCAT 2400
TTTTAGATGT TATACTTGAA ATACTGCATA ACTTTTAGCT TTCATGGGTT CCTTTTTTTC 2460
AGCCTTTAGG AGACTGTTAA GCAATTTGCT GTCCAACTTT TGTGTTGGTC TTAAACTGCA 2520
ATAGTAGTTT ACCTTGTATT GAAGAAATAA AGACCATTTT TATATTAAAA AATACTTTTG 2580
TCTGTCTTCA TTTTGACTTG TCTGATATCC TTGCAGTGCC CATTATGTCA GTTCTGTCAG 2640
ATATTCAGAC ATCAAAACTT AACGTGAGCT CAGTGGAGTT ACAGCTGCGG TTTTGATGCT 2700

FIG. 1A

GTTATTATTT CTGAAACTAG AAATGATGTT GTCTTCATCT GCTCATCAAA CACTTCATGC 2760
AGAGTGTAAG GCTAGTGAGA AATGCATACA TTTATTGATA CTTTTTTAAA GTCAACTTTT 2820
TATCAGATTT TTTTTTCATT TGGAAATATA TTGTTTTCTA GACTGCATAG CTTCTGAATC 2880
TGAAATGCAG TCTGATTGGC ATGAAGAAGC ACAGCACTCT TCATCTTACT TAAACTTCAT 2940
TTTGGAATGA AGGAAGTTAA GCAAGGGCAC AGGTCCATGA AATAGAGACA GTGCGCTCAG 3000
GAGAAAGTGA ACCTGGATTT CTTTGGCTAG TGTTCTAAAT CTGTAGTGAG GAAAGTAACA 3060
CCCGATTCCT TGAAAGGGCT CCAGCTTTAA TGCTTCCAAA TTGAAGGTGG CAGGCAACTT 3120
GGCCACTGGT TATTTACTGC ATTATGTCTC AGTTTCGCAG CTAACCTGGC TTCTCCACTA 3180
TTGAGCATGG ACTATAGCCT GGCTTCAGAG GCCAGGTGAA GGTTGGGATG GGTGGAAGGA 3240
GTGCTGGGCT GTGGCTGGGG GGACTGTGGG GACTCCAAGC TGAGCTTGGG GTGGGCAGCA 3300
CAGGGAAAAG TGTGGGTAAC TATTTTTAAG TACTGTGTTG CAAACGTCTC ATCTGCAAAT 3360
ACGTAGGGTG TGTACTCTCG AAGATTAACA GTGTGGGTTC AGTAATATAT GGATGAATTC 3420
ACAGTGGAAG CATTCAAGGG TAGATCATCT AACGACACCA GATCATCAAG CTATGATTGG 3480
AAGCGGTATC AGAAGAGCGA GGAAGGTAAG CAGTCTTCAT ATGTTTTCCC TCCACGTAAA 3540
GCAGTCTGGG AAAGTAGCAC CCCTTGAGCA GAGACAAGGA AATAATTCAG GAGCATGTGC 3600
TAGGAGAACT TTCTTGCTGA ATTCTACTTG CAAGAGCTTT GATGCCTGGC TTCTGGTGCC 3660
TTCTGCAGCA CCTGCAAGGC CCAGAGCCTG TGGTGAGCTG GAGGGAAAGA TTCTGCTCAA 3720
GTCCAAGCTT CAGCAGGTCA TTGTCTTTGC TTCTTCCCCC AGCACTGTGC AGCAGAGTGG 3780
AACTGATGTC GAAGCCTCCT GTCCACTACC TGTTGCTGCA GGCAGACTGC TCTCAGAAAA 3840
AGAGAGCTAA CTCTATGCCA TAGTCTGAAG GTAAAATGGG TTTTAAAAAA GAAAACACAA 3900
AGGCAAAACC GGCTGCCCCA TGAGAAGAAA GCAGTGGTAA ACATGGTAGA AAAGGTGCAG 3960
AAGCCCCCAG GCAGTGTGAC AGGCCCCTCC TGCCACCTAG AGGCGGGAAC AAGCTTCCCT 4020
GCCTAGGGCT CTGCCCGCGA AGTGCGTGTT TCTTTGGTGG GTTTTGTTTG GCGTTTGGTT 4080
TTGAGATTTA GACACAAGGG AAGCCTGAAA GGAGGTGTTG GGCACTATTT TGGTTTGTAA 4140
AGCCTGTACT TCAAATATAT ATTTTGTGAG GGAGTGTAGC GAATTGGCCA ATTTAAAATA 4200
AAGTTGCAAG AGATTGAAGG CTGAGTAGTT GAGAGGGTAA CACGTTTAAT GAGATCTTCT 4260
GAAACTACTG CTTCTAAACA CTTGTTTGAG TGGTGAGACC TTGGATAGGT GAGTGCTCTT 4320
GTTACATGTC TGATGCACTT GCTTGTCCTT TTCCATCCAC ATCCATGCAT TCCACATCCA 4380
CGCATTTGTC ACTTATCCCA TATCTGTCAT ATCTGACATA CCTGTCTCTT CGTCACTTGG 4440
TCAGAAGAAA CAGATGTGAT AATCCCCAGC CGCCCCAAGT TTGAGAAGAT GGCAGTTGCT 4500
TCTTTCCCTT TTTCCTGCTA AGTAAGGATT TTCTCCTGGC TTTGACACCT CACGAAATAG 4560
TCTTCCTGCC TTACATTCTG GGCATTATTT CAAATATCTT TGGAGTGCGC TGCTCTCAAG 4620
TTTGTGTCTT CCTACTCTTA GAGTGAATGC TCTTAGAGTG AAAGAGAAGG AAGAGAAGAT 4680
GTTGGCCGCA GTTCTCTGAT GAACACACCT CTGAATAATG GCCAAAGGTG GGTGGGTTTC 4740
TCTGAGGAAC GGGCAGCGTT TGCCTCTGAA AGCAAGGAGC TCTGCGGAGT TGCAGTTATT 4800
TTGCAACTGA TGGTGGAACT GGTGCTTAAA GCAGATTCCC TAGGTTCCCT GCTACTTCTT 4860
TTCCTTCTTG GCAGTCAGTT TATTTCTGAC AGACAAACAG CCACCCCCAC TGCAGGCTTA 4920
GAAAGTATGT GGCTCTGCCT GGGTGTGTTA CAGCTCTGCC CTGGTGAAAG GGGATTAAAA 4980
CGGGCACCAT TCATCCCAAA CAGGATCCTC ATTCATGGAT CAAGCTGTAA GGAACTTGGG 5040
CTCCAACCTC AAAACATTAA TTGGAGTACG AATGTAATTA AAACTGCATT CTCGCATTCC 5100
TAAGTCATTT AGTCTGGACT CTGCAGCATG TAGGTCGGCA GCTCCCACTT TCTCAAAGAC 5160
CACTGATGGA GGAGTAGTAA AAATGGAGAC CGATTCAGAA CAACCAACGG AGTGTTGCCG 5220
AAGAAACTGA TGGAAATAAT GCATGAATTG TGTGGTGGAC ATTTTTTTTA AATACATAAA 5280
CTACTTCAAA TGAGGTCGGA GAAGGTCAGT GTTTTATTAG CAGCCATAAA ACCAGGTGAG 5340
CGAGTACCAT TTTTCTCTAC AAGAAAAACG ATTCTGAGCT CTGCGTAAGT ATAAGTTCTC 5400

FIG. 1B

```
CATAGCGGCT GAAGCTCCCC CCTGGCTGCC TGCCATCTCA GCTGGAGTGC AGTGCCATTT 5460
CCTTGGGGTT TCTCTCACAG CAGTAATGGG ACAATACTTC ACAAAAATTC TTTCTTTTCC 5520
TGTCATGTGG GATCCCTACT GTGCCCTCCT GGTTTTACGT TACCCCCTGA CTGTTCCATT 5580
CAGCGGTTTG GAAAGAGAAA AAGAATTTGG AAATAAAACA TGTCTACGTT ATCACCTCCT 5640
CCAGCATTTT GGTTTTTAAT TATGTCAATA ACTGGCTTAG ATTTGGAAAT GAGAGGGGGT 5700
TGGGTGTATT ACCGAGGAAC AAAGGAAGGC TTATATAAAC TCAAGTCTTT TATTTAGAGA 5760
ACTGGCAAGC TGTCAAAAAC AAAAAGGCCT TACCACCAAA TTAAGTGAAT AGCCGCTATA 5820
GCCAGCAGGG CCAGCACGAG GGATGGTGCA CTGCTGGCAC TATGCCACGG CCTGCTTGTG 5880
ACTCTGAGAG CAACTGCTTT GGAAATGACA GCACTTGGTG CAATTTCCTT TGTTTCAGAA 5940
TGCGTAGAGC GTGTGCTTGG CGACAGTTTT TCTAGTTAGG CCACTTCTTT TTTCCTTCTC 6000
TCCTCATTCT CCTAAGCATG TCTCCATGCT GGTAATCCCA GTCAAGTGAA CGTTCAAACA 6060
ATGAATCCAT CACTGTAGGA TTCTCGTGGT GATCAAATCT TTGTGTGAGG TCTATAAAAT 6120
ATGGAAGCTT ATTTATTTTT CGTTCTTCCA TATCAGTCTT CTCTATGACA ATTCACATCC 6180
ACCACAGCAA ATTAAAGGTG AAGGAGGCTG GTGGGATGAA GAGGGTCTTC TAGCTTTACG 6240
TTCTTCCTTG CAAGGCCACA GGAAAATGCT GAGAGCTGTA GAATACAGCC TGGGGTAAGA 6300
AGTTCAGTCT CCTGCTGGGA CAGCTAACCG CATCTTATAA CCCCTTCTGA GACTCATCTT 6360
AGGACCAAAT AGGGTCTATC TGGGGTTTTT GTTCCTGCTG TTCCTCCTGG AAGGCTATCT 6420
CACTATTTCA CTGCTCCCAC GGTTACAAAC CAAAGATACA GCCTGAATTT TTTCTAGGCC 6480
ACATTACATA AATTTGACCT GGTACCAATA TTGTTCTCTA TATAGTTATT TCCTTCCCCA 6540
CTGTGTTTAA CCCCTTAAGG CATTCAGAAC AACTAGAATC ATAGAATGGT TTGGATTGGA 6600
AGGGGCCTTA AACATCATCC ATTTCCAACC CTCTGCCATG GGCTGCTTGC CACCCACTGG 6660
CTCAGGCTGC CCAGGGCCCC ATCCAGCCTG GCCTTGAGCA CCTCCAGGGA TGGGGCACCC 6720
ACAGCTTCTC TGGGCAGCCT GTGCCAACAC CTCACCACTC TCTGGGTAAA GAATTCTCTT 6780
TTAACATCTA ATCTAAATCT CTTCTCTTTT AGTTTAAAGC CATTCCTCTT TTTCCCGTTG 6840
CTATCTGTCC AAGAAATGTG TATTGGTCTC CCTCCTGCTT ATAAGCAGGA AGTACTGGAA 6900
GGCTGCAGTG AGGTCTCCCC ACAGCCTTCT CTTCTCCAGG CTGAACAAGC CCAGCTCCTT 6960
CAGCCTGTCT TCGTAGGAGA TCATCTTAGT GGCCCTCCTC TGGACCCATT CCAACAGTTC 7020
CACGGCTTTC TTGTGGAGCC CCAGGTCTGG ATGCAGTACT TCAGATGGGG CCTTACAAAG 7080
GCAGAGCAGA TGGGGACAAT CGCTTACCCC TCCCTGCTGG CTGCCCCTGT TTTGATGCAG 7140
CCCAGGGTAC TGTTGGCCTT TCAGGCTCCC AGACCCCTTG CTGATTTGTG TCAAGCTTTT 7200
CATCCACCAG AACCCACGCT TCCTGGTTAA TACTTCTGCC CTCACTTCTG TAAGCTTGTT 7260
TCAGGAGACT TCCATTCTTT AGGACAGACT GTGTTACACC TACCTGCCCT ATTCTTGCAT 7320
ATATACATTT CAGTTCATGT TTCCTGTAAC AGGACAGAAT ATGTATTCCT CTAACAAAAA 7380
TACATGCAGA ATTCCTAGTG CCATCTCAGT AGGGTTTTCA TGGCAGTATT AGCACATAGT 7440
CAATTTGCTG CAAGTACCTT CCAAGCTGCG GCCTCCCATA AATCCTGTAT TTGGGATCAG 7500
TTACCTTTTG GGGTAAGCTT TTGTATCTGC AGAGACCCTG GGGGTTCTGA TGTGCTTCAG 7560
CTCTGCTCTG TTCTGACTGC ACCATTTTCT AGATCACCCA GTTGTTCCTG TACAACTTCC 7620
TTGTCCTCCA TCCTTTCCCA GCTTGTATCT TTGACAAATA CAGGCCTATT TTTGTGTTTG 7680
CTTCAGCAGC CATTTAATTC TTCAGTGTCA TCTTGTTCTG TTGATGCCAC TGGAACAGGA 7740
TTTTCAGCAG TCTTGCAAAG AACATCTAGC TGAAAACTTT CTGCCATTCA ATATTCTTAC 7800
CAGTTCTTCT TGTTTGAGGT GAGCCATAAA TTACTAGAAC TTCGTCACTG ACAAGTTTAT 7860
GCATTTTATT ACTTCTATTA TGTACTTACT TTGACATAAC ACAGACACGC ACATATTTTG 7920
CTGGGATTTC CACAGTGTCT CTGTGTCCTT CACATGGTTT TACTGTCATA CTTCCGTTAT 7980
AACCTTGGCA ATCTGCCCAG CTGCCCATCA CAAGAAAAGA GATTCCTTTT TTATTACTTC 8040
```

FIG. 1C

```
TCTTCAGCCA ATAAACAAAA TGTGAGAAGC CCAAACAAGA ACTTGTGGGG CAGGCTGCCA 8100
TCAAGGGAGA GACAGCTGAA GGGTTGTGTA GCTCAATAGA ATTAAGAAAT AATAAAGCTG 8160
TGTCAGACAG TTTTGCCTGA TTTATACAGG CACGCCCCAA GCCAGAGAGG CTGTCTGCCA 8220
AGGCCACCTT GCAGTCCTTG GTTTGTAAGA TAAGTCATAG GTAACTTTTC TGGTGAATTG 8280
CGTGGAGAAT CATGATGGCA GTTCTTGCTG TTTACTATGG TAAGATGCTA AAATAGGAGA 8340
CAGCAAAGTA ACACTTGCTG CTGTAGGTGC TCTGCTATCC AGACAGCGAT GGCACTCGCA 8400
CACCAAGATG AGGGATGCTC CCAGCTGACG GATGCTGGGG CAGTAACAGT GGGTCCCATG 8460
CTGCCTGCTC ATTAGCATCA CCTCAGCCCT CACCAGCCCA TCAGAAGGAT CATCCCAAGC 8520
TGAGGAAAGT TGCTCATCTT CTTCACATCA TCAAACCTTT GGCCTGACTG ATGCCTCCCG 8580
GATGCTTAAA TGTGGTCACT GACATCTTTA TTTTTCTATG ATTTCAAGTC AGAACCTCCG 8640
GATCAGGAGG GAACACATAG TGGGAATGTA CCCTCAGCTC CAAGGCCAGA TCTTCCTTCA 8700
ATGATCATGC ATGCTACTTA GGAAGGTGTG TGTGTGTGAA TGTAGAATTG CCTTTGTTAT 8760
TTTTTCTTCC TGCTGTCAGG AACATTTTGA ATACCAGAGA AAAAGAAAAG TGCTCTTCTT 8820
GGCATGGGAG GAGTTGTCAC ACTTGCAAAA TAAAGGATGC AGTCCCAAAT GTTCATAATC 8880
TCAGGGTCTG AAGGAGGATC AGAAACTGTG TATACAATTT CAGGCTTCTC TGAATGCAGC 8940
TTTTGAAAGC TGTTCCTGGC CGAGGCAGTA CTAGTCAGAA CCCTCGGAAA CAGGAACAAA 9000
TGTCTTCAAG GTGCAGCAGG AGGAAACACC TTGCCCATCA TGAAAGTGAA TAACCACTGC 9060
CGCTGAAGGA ATCCAGCTCC TGTTTGAGCA GGTGCTGCAC ACTCCCACAC TGAAACAACA 9120
GTTCATTTTT ATAGGACTTC CAGGAAGGAT CTTCTTCTTA AGCTTCTTAA TTATGGTACA 9180
TCTCCAGTTG GCAGATGACT ATGACTACTG ACAGGAGAAT GAGGAACTAG CTGGGAATAT 9240
TTCTGTTTGA CCACCATGGA GTCACCCATT TCTTTACTGG TATTTGGAAA TAATAATTCT 9300
GAATTGCAAA GCAGGAGTTA GCGAAGATCT TCATTTCTTC CATGTTGGTG ACAGCACAGT 9360
TCTGGCTATG AAAGTCTGCT TACAAGGAAG AGGATAAAAA TCATAGGGAT AATAAATCTA 9420
AGTTTGAAGA CAATGAGGTT TTAGCTGCAT TTGACATGAA GAAATTGAGA CCTCTACTGG 9480
ATAGCTATGG TATTTACGTG TCTTTTTGCT TAGTTACTTA TTGACCCCAG CTGAGGTCAA 9540
GTATGAACTC AGGTCTCTCG GGCTACTGGC ATGGATTGAT TACATACAAC TGTAATTTTA 9600
GCAGTGATTT AGGGTTTATG AGTACTTTTG CAGTAAATCA TAGGGTTAGT AATGTTAATC 9660
TCAGGGAAAA AAAAAAAAAG CCAACCCTGA CAGACATCCC AGCTCAGGTG GAAATCAAGG 9720
ATCACAGCTC AGTGCGGTCC CAGAGAACAC AGGGACTCTT CTCTTAGGAC CTTTATGTAC 9780
AGGGCCTCAA GATAACTGAT GTTAGTCAGA AGACTTTCCA TTCTGGCCAC AGTTCAGCTG 9840
AGGCAATCCT GGAATTTTCT CTCCGCTGCA CAGTTCCAGT CATCCCAGTT TGTACAGTTC 9900
TGGCACTTTT TGGGTCAGGC CGTGATCCAA GGAGCAGAAG TTCCAGCTAT GGTCAGGGAG 9960
TGCCTGACCG TCCCAACTCA CTGCACTCAA ACAAAGGCGA AACCACAAGA GTGGCTTTTG 10020
TTGAAATTGC AGTGTGGCCC AGAGGGGCTG CACCAGTACT GGATTGACCA CGAGGCAACA 10080
TTAATCCTCA GCAAGTGCAA TTTGCAGCCA TTAAATTGAA CTAACTGATA CTACAATGCA 10140
ATCAGTATCA ACAAGTGGTT TGGCTTGGAA GATGGAGTCT AGGGGCTCTA CAGGAGTAGC 10200
TACTCTCTAA TGGAGTTGCA TTTTGAAGCA GGACACTGTG AAAAGCTGGC CTCCTAAAGA 10260
GGCTGCTAAA CATTAGGGTC AATTTTCCAG TGCACTTTCT GAAGTGTCTG CAGTTCCCCA 10320
TGCAAAGCTG CCCAAACATA GCACTTCCAA TTGAATACAA TTATATGCAG GCGTACTGCT 10380
TCTTGCCAGC ACTGTCCTTC TCAAATGAAC TCAACAAACA ATTTCAAAGT CTAGTAGAAA 10440
GTAACAAGCT TTGAATGTCA TTAAAAAGTA TATCTGCTTT CAGTAGTTCA GCTTATTTAT 10500
GCCCACTAGA AACATCTTGT ACAAGCTGAA CACTGGGGCT CCAGATTAGT GGTAAAACCT 10560
ACTTTATACA ATCATAGAAT CATAGAATGG CCTGGGTTGG AAGGGACCCC AAGGATCATG 10620
AAGATCCAAC ACCCCCGCCA CAGGCAGGGC CACCAACCTC CAGATCTGGT ACTAGACCAG 10680
GCAGCCCAGG GCTCCATCCA ACCTGGCCAT GAACACCTCC AGGGATGGAG CATCCACAAC 10740
```

FIG. 1D

```
CTCTCTGGGC AGCCTGTGCC AGCACCTCAC CACCCTCTCT GTGAAGAACT TTTCCCTGAC 10800
ATCCAATCTA AGCCTTCCCT CCTTGAGGTT AGATCCACTC CCCCTTGTGC TATCACTGTC 10860
TACTCTTGTA AAAAGTTGAT TCTCCTCCTT TTTGGAAGGT TGCAATGAGG TCTCCTTGCA 10920
GCCTTCTTCT CTTCTGCAGG ATGAACAAGC CCAGCTCCCT CAGCCTGTCT TTATAGGAGA 10980
GGTGCTCCAG CCCTCTGATC ATCTTTGTGG CCCTCCTCTG GACCCGCTCC AAGAGCTCCA 11040
CATCTTTCCT GTACTGGGGG CCCCAGGCCT GAATGCAGTA CTCCAGATGG GGCCTCAAAA 11100
GAGCAGAGTA AAGAGGGACA ATCACCTTCC TCACCCTGCT GGCCAGCCCT CTTCTGATGG 11160
AGCCCTGGAT ACAACTGGCT TTCTGAGCTG CAACTTCTCC TTATCAGTTC CACTATTAAA 11220
ACAGGAACAA TACAACAGGT GCTGATGGCC AGTGCAGAGT TTTTCACACT TCTTCATTTC 11280
GGTAGATCTT AGATGAGGAA CGTTGAAGTT GTGCTTCTGC GTGTGCTTCT TCCTCCTCAA 11340
ATACTCCTGC CTGATACCTC ACCCCACCTG CCACTGAATG GCTCCATGGC CCCCTGCAGC 11400
CAGGGCCCTG ATGAACCCGG CACTGCTTCA GATGCTGTTT AATAGCACAG TATGACCAAG 11460
TTGCACCTAT GAATACACAA ACAATGTGTT GCATCCTTCA GCACTTGAGA AGAAGAGCCA 11520
AATTTGCATT GTCAGGAAAT GGTTTAGTAA TTCTGCCAAT TAAAACTTGT TTATCTACCA 11580
TGGCTGTTTT TATGGCTGTT AGTAGTGGTA CACTGATGAT GAACAATGGC TATGCAGTAA 11640
AATCAAGACT GTAGATATTG CAACAGACTA TAAAATTCCT CTGTGGCTTA GCCAATGTGG 11700
TACTTCCCAC ATTGTATAAG AAATTTGGCA AGTTTAGAGC AATGTTTGAA GTGTTGGGAA 11760
ATTTCTGTAT ACTCAAGAGG GCGTTTTTGA CAACTGTAGA ACAGAGGAAT CAAAAGGGGG 11820
TGGGAGGAAG TTAAAAGAAG AGGCAGGTGC AAGAGAGCTT GCAGTCCCGC TGTGTGTACG 11880
ACACTGGCAA CATGAGGTCT TTGCTAATCT TGGTGCTTTG CTTCCTGCCC CTGGCTGCCT 11940
TAGGGTGCGA TCTGCCTCAG ACCCACAGCC TGGGCAGCAG GAGGACCCTG ATGCTGCTGG 12000
CTCAGATGAG GAGAATCAGC CTGTTTAGCT GCCTGAAGGA TAGGCACGAT TTTGGCTTTC 12060
CTCAAGAGGA GTTTGGCAAC CAGTTTCAGA AGGCTGAGAC CATCCCTGTG CTGCACGAGA 12120
TGATCCAGCA GATCTTTAAC CTGTTTAGCA CCAAGGATAG CAGCGCTGCT TGGGATGAGA 12180
CCCTGCTGGA TAAGTTTTAC ACCGAGCTGT ACCAGCAGCT GAACGATCTG GAGGCTTGCG 12240
TGATCCAGGG CGTGGGCGTG ACCGAGACCC CTCTGATGAA GGAGGATAGC ATCCTGGCTG 12300
TGAGGAAGTA CTTTCAGAGG ATCACCCTGT ACCTGAAGGA GAAGAAGTAC AGCCCCTGCG 12360
CTTGGGAAGT CGTGAGGGCT GAGATCATGA GGAGCTTTAG CCTGAGCACC AACCTGCAAG 12420
AGAGCTTGAG GTCTAAGGAG TAAAAAGTCT AGAGTCGGGG CGGCCGGCCG CTTCGAGCAG 12480
ACATGATAAG ATACATTGAT GAGTTTGGAC AAACCACAAC TAGAATGCAG TGAAAAAAAT 12540
GCTTTATTTG TGAAATTTGT GATGCTATTG CTTTATTTGT AACCATTATA AGCTGCAATA 12600
AACAAGTTAA CAACAACAAT TGCATTCATT TTATGTTTCA GGTTCAGGGG GAGGTGTGGG 12660
AGGTTTTTTA AAGCAAGTAA AACCTCTACA AATGTGGTAA AATCGATAAG GATCCGTCGA 12720
GCGGCCGC        12728
```

FIG. 1E

SEQ ID NO: 5

```
TGCGATCTGC CTCAGACCCA CAGCCTGGGC AGCAGGAGGA CCCTGATGCT GCTGGCTCAG  60
ATGAGGAGAA TCAGCCTGTT TAGCTGCCTG AAGGATAGGC ACGATTTTGG CTTTCCTCAA 120
GAGGAGTTTG GCAACCAGTT TCAGAAGGCT GAGACCATCC CTGTGCTGCA CGAGATGATC 180
CAGCAGATCT TTAACCTGTT TAGCACCAAG GATAGCAGCG CTGCTTGGGA TGAGACCCTG 240
CTGGATAAGT TTTACACCGA GCTGTACCAG CAGCTGAACG ATCTGGAGGC TTGCGTGATC 300
CAGGGCGTGG GCGTGACCGA GACCCCTCTG ATGAAGGAGG ATAGCATCCT GGCTGTGAGG 360
AAGTACTTTC AGAGGATCAC CCTGTACCTG AAGGAGAAGA AGTACAGCCC CTGCGCTTGG 420
GAAGTCGTGA GGGCTGAGAT CATGAGGAGC TTTAGCCTGA GCACCAACCT GCAAGAGAGC 480
TTGAGGTCTA AGGAGTAA 498
```

FIG. 2

SEQ ID NO: 7

TGCCGCCTTC TTTGATATTC ACTCTGTTGT ATTTCATCTC TTCTTGCCGA TGAAAGGATA 60
TAACAGTCTG TATAACAGTC TGTGAGGAAA TACTTGGTAT TTCTTCTGAT CAGTGTTTTT 120
ATAAGTAATG TTGAATATTG GATAAGGCTG TGTGTCCTTT GTCTTGGGAG ACAAAGCCCA 180
CAGCAGGTGG TGGTTGGGGT GGTGGCAGCT CAGTGACAGG AGAGGTTTTT TTGCCTGTTT 240
TTTTTTTTTT TTTTTTTTTT AAGTAAGGTG TTCTTTTTTC TTAGTAAATT TTCTACTGGA 300
CTGTATGTTT TGACAGGTCA GAAACATTTC TTCAAAAGAA GAACCTTTTG GAAACTGTAC 360
AGCCCTTTTC TTTCATTCCC TTTTTGCTTT CTGTGCCAAT GCCTTTGGTT CTGATTGCAT 420
TATGGAAAAC GTTGATCGGA ACTTGAGGTT TTTATTTATA GTGTGGCTTG AAAGCTTGGA 480
TAGCTGTTGT TACACGAGAT ACCTTATTAA GTTTAGGCCA GCTTGATGCT TTATTTTTTC 540
CCTTTGAAGT AGTGAGCGTT CTCTGGTTTT TTTCCTTTGA AACTGGTGAG GCTTAGATTT 600
TTCTAATGGG ATTTTTTACC TGATGATCTA GTTGCATACC CAAATGCTTG TAAATGTTTT 660
CCTAGTTAAC ATGTTGATAA CTTCGGATTT ACATGTTGTA TATACTTGTC ATCTGTGTTT 720
CTAGTAAAAA TATATGGCAT TTATAGAAAT ACGTAATTCC TGATTTCCTT TTTTTTTATC 780
TCTATGCTCT GTGTGTACAG GTCAAACAGA CTTCACTCCT ATTTTTATTT ATAGAATTTT 840
ATATGCAGTC TGTCGTTGGT TCTTGTGTTG TAAGGATACA GCCTTAAATT TCCTAGAGCG 900
ATGCTCAGTA AGGCGGGTTG TCACATGGGT TCAAATGTAA AACGGGCACG TTTGGCTGCT 960
GCCTTCCCGA GATCCAGGAC ACTAAACTGC TTCTGCACTG AGGTATAAAT CGCTTCAGAT 1020
CCCAGGGAAG TGCAGATCCA CGTGCATATT CTTAAAGAAG AATGAATACT TTCTAAAATA 1080
TTTTGGCATA GGAAGCAAGC TGCATGGATT TGTTTGGGAC TTAAATTATT TTGGTAACGG 1140
AGTGCATAGG TTTTAAACAC AGTTGCAGCA TGCTAACGAG TCACAGCGTT TATGCAGAAG 1200
TGATGCCTGG ATGCCTGTTG CAGCTGTTTA CGGCACTGCC TTGCAGTGAG CATTGCAGAT 1260
AGGGGTGGGG TGCTTTGTGT CGTGTTCCCA CACGCTGCCA CACAGCCACC TCCCGGAACA 1320
CATCTCACCT GCTGGGTACT TTTCAAACCA TCTTAGCAGT AGTAGATGAG TTACTATGAA 1380
ACAGAGAAGT TCCTCAGTTG GATATTCTCA TGGGATGTCT TTTTTCCCAT GTTGGGCAAA 1440
GTATGATAAA GCATCTCTAT TTGTAAATTA TGCACTTGTT AGTTCCTGAA TCCTTTCTAT 1500
AGCACCACTT ATTGCAGCAG GTGTAGGCTC TGGTGTGGCC TGTGTCTGTG CTTCAATCTT 1560
TTAAAGCTTC TTTGGAAATA CACTGACTTG ATTGAAGTCT CTTGAAGATA GTAAACAGTA 1620
CTTACCTTTG ATCCCAATGA AATCGAGCAT TTCAGTTGTA AAAGAATTCC GCCTATTCAT 1680
ACCATGTAAT GTAATTTTAC ACCCCCAGTG CTGACACTTT GGAATATATT CAAGTAATAG 1740
ACTTTGGCCT CACCCTCTTG TGTACTGTAT TTTGTAATAG AAAATATTTT AAACTGTGCA 1800
TATGATTATT ACATTATGAA AGAGACATTC TGCTGATCTT CAAATGTAAG AAAATGAGGA 1860
GTGCGTGTGC TTTTATAAAT ACAAGTGATT GCAAATTAGT GCAGGTGTCC TTAAAAAAAA 1920
AAAAAAAAAG TAATATAAAA AGGACCAGGT GTTTTACAAG TGAAATACAT TCCTATTTGG 1980
TAAACAGTTA CATTTTTATG AAGATTACCA GCGCTGCTGA CTTTCTAAAC ATAAGGCTGT 2040
ATTGTCTTCC TGTACCATTG CATTTCCTCA TTCCCAATTT GCACAAGGAT GTCTGGGTAA 2100
ACTATTCAAG AAATGGCTTT GAAATACAGC ATGGGAGCTT GTCTGAGTTG GAATGCAGAG 2160
TTGCACTGCA AAATGTCAGG AAATGGATGT CTCTCAGAAT GCCCAACTCC AAAGGATTTT 2220
ATATGTGTAT ATAGTAAGCA GTTTCCTGAT TCCAGCAGGC CAAAGAGTCT GCTGAATGTT 2280
GTGTTGCCGG AGACCTGTAT TTCTCAACAA GGTAAGATGG TATCCTAGCA ACTGCGGATT 2340
TTAATACATT TTCAGCAGAA GTACTTAGTT AATCTCTACC TTTAGGGATC GTTTCATCAT 2400
TTTTAGATGT TATACTTGAA ATACTGCATA ACTTTTAGCT TTCATGGGTT CCTTTTTTTC 2460
AGCCTTTAGG AGACTGTTAA GCAATTTGCT GTCCAACTTT TGTGTTGGTC TTAAACTGCA 2520
ATAGTAGTTT ACCTTGTATT GAAGAAATAA AGACCATTTT TATATTAAAA AATACTTTTG 2580
TCTGTCTTCA TTTTGACTTG TCTGATATCC TTGCAGTGCC CATTATGTCA GTTCTGTCAG 2640
ATATTCAGAC ATCAAAACTT AACGTGAGCT CAGTGGAGTT ACAGCTGCGG TTTTGATGCT 2700

FIG. 3A

```
GTTATTATTT CTGAAACTAG AAATGATGTT GTCTTCATCT GCTCATCAAA CACTTCATGC 2760
AGAGTGTAAG GCTAGTGAGA AATGCATACA TTTATTGATA CTTTTTTAAA GTCAACTTTT 2820
TATCAGATTT TTTTTTCATT TGGAAATATA TTGTTTTCTA GACTGCATAG CTTCTGAATC 2880
TGAAATGCAG TCTGATTGGC ATGAAGAAGC ACAGCACTCT TCATCTTACT TAAACTTCAT 2940
TTTGGAATGA AGGAAGTTAA GCAAGGGCAC AGGTCCATGA AATAGAGACA GTGCGCTCAG 3000
GAGAAAGTGA ACCTGGATTT CTTTGGCTAG TGTTCTAAAT CTGTAGTGAG GAAAGTAACA 3060
CCCGATTCCT TGAAAGGGCT CCAGCTTTAA TGCTTCCAAA TTGAAGGTGG CAGGCAACTT 3120
GGCCACTGGT TATTTACTGC ATTATGTCTC AGTTTCGCAG CTAACCTGGC TTCTCCACTA 3180
TTGAGCATGG ACTATAGCCT GGCTTCAGAG GCCAGGTGAA GGTTGGGATG GGTGGAAGGA 3240
GTGCTGGGCT GTGGCTGGGG GGACTGTGGG GACTCCAAGC TGAGCTTGGG GTGGGCAGCA 3300
CAGGGAAAAG TGTGGGTAAC TATTTTTAAG TACTGTGTTG CAAACGTCTC ATCTGCAAAT 3360
ACGTAGGGTG TGTACTCTCG AAGATTAACA GTGTGGGTTC AGTAATATAT GGATGAATTC 3420
ACAGTGGAAG CATTCAAGGG TAGATCATCT AACGACACCA GATCATCAAG CTATGATTGG 3480
AAGCGGTATC AGAAGAGCGA GGAAGGTAAG CAGTCTTCAT ATGTTTTCCC TCCACGTAAA 3540
GCAGTCTGGG AAAGTAGCAC CCCTTGAGCA GAGACAAGGA AATAATTCAG GAGCATGTGC 3600
TAGGAGAACT TTCTTGCTGA ATTCTACTTG CAAGAGCTTT GATGCCTGGC TTCTGGTGCC 3660
TTCTGCAGCA CCTGCAAGGC CCAGAGCCTG TGGTGAGCTG GAGGGAAAGA TTCTGCTCAA 3720
GTCCAAGCTT CAGCAGGTCA TTGTCTTTGC TTCTTCCCCC AGCACTGTGC AGCAGAGTGG 3780
AACTGATGTC GAAGCCTCCT GTCCACTACC TGTTGCTGCA GGCAGACTGC TCTCAGAAAA 3840
AGAGAGCTAA CTCTATGCCA TAGTCTGAAG GTAAAATGGG TTTTAAAAAA GAAAACACAA 3900
AGGCAAAACC GGCTGCCCCA TGAGAAGAAA GCAGTGGTAA ACATGGTAGA AAAGGTGCAG 3960
AAGCCCCCAG GCAGTGTGAC AGGCCCCTCC TGCCACCTAG AGGCGGGAAC AAGCTTCCCT 4020
GCCTAGGGCT CTGCCCGCGA AGTGCGTGTT TCTTTGGTGG GTTTTGTTTG GCGTTTGGTT 4080
TTGAGATTTA GACACAAGGG AAGCCTGAAA GGAGGTGTTG GGCACTATTT TGGTTTGTAA 4140
AGCCTGTACT TCAAATATAT ATTTTGTGAG GGAGTGTAGC GAATTGGCCA ATTTAAAATA 4200
AAGTTGCAAG AGATTGAAGG CTGAGTAGTT GAGAGGGTAA CACGTTTAAT GAGATCTTCT 4260
GAAACTACTG CTTCTAAACA CTTGTTTGAG TGGTGAGACC TTGGATAGGT GAGTGCTCTT 4320
GTTACATGTC TGATGCACTT GCTTGTCCTT TTCCATCCAC ATCCATGCAT TCCACATCCA 4380
CGCATTTGTC ACTTATCCCA TATCTGTCAT ATCTGACATA CCTGTCTCTT CGTCACTTGG 4440
TCAGAAGAAA CAGATGTGAT AATCCCCAGC CGCCCCAAGT TTGAGAAGAT GGCAGTTGCT 4500
TCTTTCCCTT TTTCCTGCTA AGTAAGGATT TTCTCCTGGC TTTGACACCT CACGAAATAG 4560
TCTTCCTGCC TTACATTCTG GGCATTATTT CAAATATCTT TGGAGTGCGC TGCTCTCAAG 4620
TTTGTGTCTT CCTACTCTTA GAGTGAATGC TCTTAGAGTG AAAGAGAAGG AAGAGAAGAT 4680
GTTGGCCGCA GTTCTCTGAT GAACACACCT CTGAATAATG GCCAAAGGTG GGTGGGTTTC 4740
TCTGAGGAAC GGGCAGCGTT TGCCTCTGAA AGCAAGGAGC TCTGCGGAGT TGCAGTTATT 4800
TTGCAACTGA TGGTGGAACT GGTGCTTAAA GCAGATTCCC TAGGTTCCCT GCTACTTCTT 4860
TTCCTTCTTG GCAGTCAGTT TATTTCTGAC AGACAAACAG CCACCCCCAC TGCAGGCTTA 4920
GAAAGTATGT GGCTCTGCCT GGGTGTGTTA CAGCTCTGCC CTGGTGAAAG GGGATTAAAA 4980
CGGGCACCAT TCATCCCAAA CAGGATCCTC ATTCATGGAT CAAGCTGTAA GGAACTTGGG 5040
CTCCAACCTC AAAACATTAA TTGGAGTACG AATGTAATTA AAACTGCATT CTCGCATTCC 5100
TAAGTCATTT AGTCTGGACT CTGCAGCATG TAGGTCGGCA GCTCCCACTT TCTCAAAGAC 5160
CACTGATGGA GGAGTAGTAA AAATGGAGAC CGATTCAGAA CAACCAACGG AGTGTTGCCG 5220
AAGAAACTGA TGGAAATAAT GCATGAATTG TGTGGTGGAC ATTTTTTTTA AATACATAAA 5280
CTACTTCAAA TGAGGTCGGA GAAGGTCAGT GTTTTATTAG CAGCCATAAA ACCAGGTGAG 5340
CGAGTACCAT TTTTCTCTAC AAGAAAAACG ATTCTGAGCT CTGCGTAAGT ATAAGTTCTC 5400
```

FIG. 3B

```
CATAGCGGCT GAAGCTCCCC CCTGGCTGCC TGCCATCTCA GCTGGAGTGC AGTGCCATTT 5460
CCTTGGGGTT TCTCTCACAG CAGTAATGGG ACAATACTTC ACAAAAATTC TTTCTTTTCC 5520
TGTCATGTGG GATCCCTACT GTGCCCTCCT GGTTTTACGT TACCCCCTGA CTGTTCCATT 5580
CAGCGGTTTG GAAAGAGAAA AAGAATTTGG AAATAAAACA TGTCTACGTT ATCACCTCCT 5640
CCAGCATTTT GGTTTTTAAT TATGTCAATA ACTGGCTTAG ATTTGGAAAT GAGAGGGGGT 5700
TGGGTGTATT ACCGAGGAAC AAAGGAAGGC TTATATAAAC TCAAGTCTTT TATTTAGAGA 5760
ACTGGCAAGC TGTCAAAAAC AAAAAGGCCT TACCACCAAA TTAAGTGAAT AGCCGCTATA 5820
GCCAGCAGGG CCAGCACGAG GGATGGTGCA CTGCTGGCAC TATGCCACGG CCTGCTTGTG 5880
ACTCTGAGAG CAACTGCTTT GGAAATGACA GCACTTGGTG CAATTTCCTT TGTTTCAGAA 5940
TGCGTAGAGC GTGTGCTTGG CGACAGTTTT TCTAGTTAGG CCACTTCTTT TTTCCTTCTC 6000
TCCTCATTCT CCTAAGCATG TCTCCATGCT GGTAATCCCA GTCAAGTGAA CGTTCAAACA 6060
ATGAATCCAT CACTGTAGGA TTCTCGTGGT GATCAAATCT TTGTGTGAGG TCTATAAAAT 6120
ATGGAAGCTT ATTTATTTTT CGTTCTTCCA TATCAGTCTT CTCTATGACA ATTCACATCC 6180
ACCACAGCAA ATTAAAGGTG AAGGAGGCTG GTGGGATGAA GAGGGTCTTC TAGCTTTACG 6240
TTCTTCCTTG CAAGGCCACA GGAAAATGCT GAGAGCTGTA GAATACAGCC TGGGGTAAGA 6300
AGTTCAGTCT CCTGCTGGGA CAGCTAACCG CATCTTATAA CCCCTTCTGA GACTCATCTT 6360
AGGACCAAAT AGGGTCTATC TGGGGTTTTT GTTCCTGCTG TTCCTCCTGG AAGGCTATCT 6420
CACTATTTCA CTGCTCCCAC GGTTACAAAC CAAAGATACA GCCTGAATTT TTTCTAGGCC 6480
ACATTACATA AATTTGACCT GGTACCAATA TTGTTCTCTA TATAGTTATT TCCTTCCCCA 6540
CTGTGTTTAA CCCCTTAAGG CATTCAGAAC AACTAGAATC ATAGAATGGT TTGGATTGGA 6600
AGGGGCCTTA AACATCATCC ATTTCCAACC CTCTGCCATG GGCTGCTTGC CACCCACTGG 6660
CTCAGGCTGC CCAGGGCCCC ATCCAGCCTG GCCTTGAGCA CCTCCAGGGA TGGGGCACCC 6720
ACAGCTTCTC TGGGCAGCCT GTGCCAACAC CTCACCACTC TCTGGGTAAA GAATTCTCTT 6780
TTAACATCTA ATCTAAATCT CTTCTCTTTT AGTTTAAAGC CATTCCTCTT TTTCCCGTTG 6840
CTATCTGTCC AAGAAATGTG TATTGGTCTC CCTCCTGCTT ATAAGCAGGA AGTACTGGAA 6900
GGCTGCAGTG AGGTCTCCCC ACAGCCTTCT CTTCTCCAGG CTGAACAAGC CCAGCTCCTT 6960
CAGCCTGTCT TCGTAGGAGA TCATCTTAGT GGCCCTCCTC TGGACCCATT CCAACAGTTC 7020
CACGGCTTTC TTGTGGAGCC CCAGGTCTGG ATGCAGTACT TCAGATGGGG CCTTACAAAG 7080
GCAGAGCAGA TGGGGACAAT CGCTTACCCC TCCCTGCTGG CTGCCCCTGT TTTGATGCAG 7140
CCCAGGGTAC TGTTGGCCTT TCAGGCTCCC AGACCCCTTG CTGATTTGTG TCAAGCTTTT 7200
CATCCACCAG AACCCACGCT TCCTGGTTAA TACTTCTGCC CTCACTTCTG TAAGCTTGTT 7260
TCAGGAGACT TCCATTCTTT AGGACAGACT GTGTTACACC TACCTGCCCT ATTCTTGCAT 7320
ATATACATTT CAGTTCATGT TTCCTGTAAC AGGACAGAAT ATGTATTCCT CTAACAAAAA 7380
TACATGCAGA ATTCCTAGTG CCATCTCAGT AGGGTTTTCA TGGCAGTATT AGCACATAGT 7440
CAATTTGCTG CAAGTACCTT CCAAGCTGCG GCCTCCCATA AATCCTGTAT TTGGGATCAG 7500
TTACCTTTTG GGGTAAGCTT TTGTATCTGC AGAGACCCTG GGGGTTCTGA TGTGCTTCAG 7560
CTCTGCTCTG TTCTGACTGC ACCATTTTCT AGATCACCCA GTTGTTCCTG TACAACTTCC 7620
TTGTCCTCCA TCCTTTCCCA GCTTGTATCT TTGACAAATA CAGGCCTATT TTTGTGTTTG 7680
CTTCAGCAGC CATTTAATTC TTCAGTGTCA TCTTGTTCTG TTGATGCCAC TGGAACAGGA 7740
TTTTCAGCAG TCTTGCAAAG AACATCTAGC TGAAAACTTT CTGCCATTCA ATATTCTTAC 7800
CAGTTCTTCT TGTTTGAGGT GAGCCATAAA TTACTAGAAC TTCGTCACTG ACAAGTTTAT 7860
GCATTTTATT ACTTCTATTA TGTACTTACT TTGACATAAC ACAGACACGC ACATATTTTG 7920
CTGGGATTTC CACAGTGTCT CTGTGTCCTT CACATGGTTT TACTGTCATA CTTCCGTTAT 7980
AACCTTGGCA ATCTGCCCAG CTGCCCATCA CAAGAAAAGA GATTCCTTTT TTATTACTTC 8040
TCTTCAGCCA ATAAACAAAA TGTGAGAAGC CCAAACAAGA ACTTGTGGGG CAGGCTGCCA 8100
```

FIG. 3C

TCAAGGGAGA GACAGCTGAA GGGTTGTGTA GCTCAATAGA ATTAAGAAAT AATAAAGCTG 8160
TGTCAGACAG TTTTGCCTGA TTTATACAGG CACGCCCCAA GCCAGAGAGG CTGCCTGCCA 8220
AGGCCACCTT GCAGTCCTTG GTTTGTAAGA TAAGTCATAG GTAACTTTTC TGGTGAATTG 8280
CGTGGAGAAT CATGATGGCA GTTCTTGCTG TTTACTATGG TAAGATGCTA AAATAGGAGA 8340
CAGCAAAGTA ACACTTGCTG CTGTAGGTGC TCTGCTATCC AGACAGCGAT GGCACTCGCA 8400
CACCAAGATG AGGGATGCTC CCAGCTGACG GATGCTGGGG CAGTAACAGT GGGTCCCATG 8460
CTGCCTGCTC ATTAGCATCA CCTCAGCCCT CACCAGCCCA TCAGAAGGAT CATCCCAAGC 8520
TGAGGAAAGT TGCTCATCTT CTTCACATCA TCAAACCTTT GGCCTGACTG ATGCCTCCCG 8580
GATGCTTAAA TGTGGTCACT GACATCTTTA TTTTTCTATG ATTTCAAGTC AGAACCTCCG 8640
GATCAGGAGG GAACACATAG TGGGAATGTA CCCTCAGCTC CAAGGCCAGA TCTTCCTTCA 8700
ATGATCATGC ATGCTACTTA GGAAGGTGTG TGTGTGTGAA TGTAGAATTG CCTTTGTTAT 8760
TTTTTCTTCC TGCTGTCAGG AACATTTTGA ATACCAGAGA AAAAGAAAAG TGCTCTTCTT 8820
GGCATGGGAG GAGTTGTCAC ACTTGCAAAA TAAAGGATGC AGTCCCAAAT GTTCATAATC 8880
TCAGGGTCTG AAGGAGGATC AGAAACTGTG TATACAATTT CAGGCTTCTC TGAATGCAGC 8940
TTTTGAAAGC TGTTCCTGGC CGAGGCAGTA CTAGTCAGAA CCCTCGGAAA CAGGAACAAA 9000
TGTCTTCAAG GTGCAGCAGG AGGAAACACC TTGCCCATCA TGAAAGTGAA TAACCACTGC 9060
CGCTGAAGGA ATCCAGCTCC TGTTTGAGCA GGTGCTGCAC ACTCCCACAC TGAAACAACA 9120
GTTCATTTTT ATAGGACTTC CAGGAAGGAT CTTCTTCTTA AGCTTCTTAA TTATGGTACA 9180
TCTCCAGTTG GCAGATGACT ATGACTACTG ACAGGAGAAT GAGGAACTAG CTGGGAATAT 9240
TTCTGTTTGA CCACCATGGA GTCACCCATT TCTTTACTGG TATTTGGAAA TAATAATTCT 9300
GAATTGCAAA GCAGGAGTTA GCGAAGATCT TCATTTCTTC CATGTTGGTG ACAGCACAGT 9360
TCTGGCTATG AAAGTCTGCT TACAAGGAAG AGGATAAAAA TCATAGGGAT AATAAATCTA 9420
AGTTTGAAGA CAATGAGGTT TTAGCTGCAT TTGACATGAA GAAATTGAGA CCTCTACTGG 9480
ATAGCTATGG TATTTACGTG TCTTTTTGCT TAGTTACTTA TTGACCCCAG CTGAGGTCAA 9540
GTATGAACTC AGGTCTCTCG GGCTACTGGC ATGGATTGAT TACATACAAC TGTAATTTTA 9600
GCAGTGATTT AGGGTTTATG AGTACTTTTG CAGTAAATCA TAGGGTTAGT AATGTTAATC 9660
TCAGGGAAAA AAAAAAAAAG CCAACCCTGA CAGACATCCC AGCTCAGGTG GAAATCAAGG 9720
ATCACAGCTC AGTGCGGTCC CAGAGAACAC AGGGACTCTT CTCTTAGGAC CTTTATGTAC 9780
AGGGCCTCAA GATAACTGAT GTTAGTCAGA AGACTTTCCA TTCTGGCCAC AGTTCAGCTG 9840
AGGCAATCCT GGAATTTTCT CTCCGCTGCA CAGTTCCAGT CATCCCAGTT TGTACAGTTC 9900
TGGCACTTTT TGGGTCAGGC CGTGATCCAA GGAGCAGAAG TTCCAGCTAT GGTCAGGGAG 9960
TGCCTGACCG TCCCAACTCA CTGCACTCAA ACAAAGGCGA AACCACAAGA GTGGCTTTTG 10020
TTGAAATTGC AGTGTGGCCC AGAGGGGCTG CACCAGTACT GGATTGACCA CGAGGCAACA 10080
TTAATCCTCA GCAAGTGCAA TTTGCAGCCA TTAAATTGAA CTAACTGATA CTACAATGCA 10140
ATCAGTATCA ACAAGTGGTT TGGCTTGGAA GATGGAGTCT AGGGGCTCTA CAGGAGTAGC 10200
TACTCTCTAA TGGAGTTGCA TTTTGAAGCA GGACACTGTG AAAAGCTGGC CTCCTAAAGA 10260
GGCTGCTAAA CATTAGGGTC AATTTTCCAG TGCACTTTCT GAAGTGTCTG CAGTTCCCCA 10320
TGCAAAGCTG CCCAAACATA GCACTTCCAA TTGAATACAA TTATATGCAG GCGTACTGCT 10380
TCTTGCCAGC ACTGTCCTTC TCAAATGAAC TCAACAAACA ATTTCAAAGT CTAGTAGAAA 10440
GTAACAAGCT TTGAATGTCA TTAAAAAGTA TATCTGCTTT CAGTAGTTCA GCTTATTTAT 10500
GCCCACTAGA AACATCTTGT ACAAGCTGAA CACTGGGGCT CCAGATTAGT GGTAAAACCT 10560
ACTTTATACA ATCATAGAAT CATAGAATGG CCTGGGTTGG AAGGGACCCC AAGGATCATG 10620
AAGATCCAAC ACCCCCGCCA CAGGCAGGGC CACCAACCTC CAGATCTGGT ACTAGACCAG 10680
GCAGCCCAGG GCTCCATCCA ACCTGGCCAT GAACACCTCC AGGGATGGAG CATCCACAAC 10740
CTCTCTGGGC AGCCTGTGCC AGCACCTCAC CACCCTCTCT GTGAAGAACT TTTCCCTGAC 10800
ATCCAATCTA AGCCTTCCCT CCTTGAGGTT AGATCCACTC CCCCTTGTGC TATCACTGTC 10860

FIG. 3D

```
TACTCTTGTA AAAAGTTGAT TCTCCTCCTT TTTGGAAGGT TGCAATGAGG TCTCCTTGCA 10920
GCCTTCTTCT CTTCTGCAGG ATGAACAAGC CCAGCTCCCT CAGCCTGTCT TTATAGGAGA 10980
GGTGCTCCAG CCCTCTGATC ATCTTTGTGG CCCTCCTCTG GACCCGCTCC AAGAGCTCCA 11040
CATCTTTCCT GTACTGGGGG CCCCAGGCCT GAATGCAGTA CTCCAGATGG GGCCTCAAAA 11100
GAGCAGAGTA AAGAGGGACA ATCACCTTCC TCACCCTGCT GGCCAGCCCT CTTCTGATGG 11160
AGCCCTGGAT ACAACTGGCT TTCTGAGCTG CAACTTCTCC TTATCAGTTC CACTATTAAA 11220
ACAGGAACAA TACAACAGGT GCTGATGGCC AGTGCAGAGT TTTTCACACT TCTTCATTTC 11280
GGTAGATCTT AGATGAGGAA CGTTGAAGTT GTGCTTCTGC GTGTGCTTCT TCCTCCTCAA 11340
ATACTCCTGC CTGATACCTC ACCCCACCTG CCACTGAATG GCTCCATGGC CCCCTGCAGC 11400
CAGGGCCCTG ATGAACCCGG CACTGCTTCA GATGCTGTTT AATAGCACAG TATGACCAAG 11460
TTGCACCTAT GAATACACAA ACAATGTGTT GCATCCTTCA GCACTTGAGA AGAAGAGCCA 11520
AATTTGCATT GTCAGGAAAT GGTTTAGTAA TTCTGCCAAT TAAAACTTGT TTATCTACCA 11580
TGGCTGTTTT TATGGCTGTT AGTAGTGGTA CACTGATGAT GAACAATGGC TATGCAGTAA 11640
AATCAAGACT GTAGATATTG CAACAGACTA TAAAATTCCT CTGTGGCTTA GCCAATGTGG 11700
TACTTCCCAC ATTGTATAAG AAATTTGGCA AGTTTAGAGC AATGTTTGAA GTGTTGGGAA 11760
ATTTCTGTAT ACTCAAGAGG GCGTTTTTGA CAACTGTAGA ACAGAGGAAT CAAAAGGGGG 11820
TGGGAGGAAG TTAAAAGAAG AGGCAGGTGC AAGAGAGCTT GCAGTCCCGC TGTGTGTACG 11880
ACACTGGCAA CATGAGGTCT TTGCTAATCT TGGTGCTTTG CTTCCTGCCC CTGGCTGCCT 11940
TAGGG  11945
```

FIG. 3E

SEQ ID NO: 8

```
AAAGTCTAGAGTCGGGGCGGCCGGCCGCTTCGAGCAGACATGATAAGATACATTGATGAG  60
TTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGAT 120
GCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGC 180
ATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAAC 240
CTCTACAAATGTGGTAAAATCGATAAGGATCCGTCGAGCGGCCGC  285
```

FIG. 4

SEQ ID NO: 9

```
   1 CGCGTGGTAGGTGGCGGGGGGTTCCCAGGAGAGCCCCCAGCGCGGACGGC
     AGCGCCGTCACTCACCGCTCCGTCTCCCTCCGCCCAGGGTCGCCTGGCGC
     AACCGCTGCAAGGGCACCGACGTCCAGGCGTGGATCAGAGGCTGCCGGCT
     GTGAGGAGCTGCCGCGCCCGGCCCGCCCGCTGCACAGCCGGCCGCTTTGC
 200 GAGCGCGACGCTACCCGCTTGGCAGTTTTAAACGCATCCCTCATTAAAAC
     GACTATACGCAAACGCCTTCCCGTCGGTCCGCGTCTCTTTCCGCCGCCAG
     GGCGACACTCGCGGGGAGGGCGGGAAGGGGGCCGGGCGGGAGCCCGCGGC
     CAACCGTCGCCCCGTGACGGCACCGCCCCGCCCCCGTGACGCGGTGCGGG
 400 CGCCGGGGCCGTGGGGCTCAGCGCTGCGGCGGGGCCGGGCCGGGCCGGGG
     CGGGAGCTGAGCGCGGCGCGGCTGCGGGCGGCGCCCCCTCCGGTGCAATA
     TGTTCAAGAGAATGGCTGAGTTCGGGCCTGACTCCGGGGGCAGGGTGAAG
     GTGCGGCGCGGGCGGAGGGACGGGGCGGGCGCGGGGCCGCCCGGCGGGTG
 600 CCGGGGCCTCTGCCGGCCCGCCCGGCTCGGGCTGCTGCGGCGCTTACGGG
     CGCGCTTCTCGCCGCTGCCGCTTCTCTTCTCTCCCGCGCAAGGGCGTCAC
     CATCGTGAAGCCGGTAGTGTACGGGAACGTGGCGCGGTACTTCGGGAAGA
     AGAGGGAGGAGGACGGGCACACGCATCAGTGGACGGTTTACGTGAAGCCC
 800 TACAGGAACGAGGTAGGGCCCGAGCGCGTCGGCCGCCGTTCTCGGAGCGC
     CGGAGCCGTCAGCGCCGCGCCTGGGTGCGCTGTGGGACACAGCGAGCTTC
     TCTCGTAGGACATGTCCGCCTACGTGAAAAAAATCCAGTTCAAGCTGCAC
     GAGAGCTACGGGAATCCTCTCCGAGGTGGGTGTTGCGTCGGGGGGTTTGC
1000 TCCGCTCGGTCCCGCTGAGGCTCGTCGCCCTCATCTTTCTTTCGTGCCGC
     AGTCGTTACCAAACCGCCGTACGAGATCACCGAAACGGGCTGGGGCGAAT
     TTGAAATCATCATCAAGATATTTTTCATTGATCCAAACGAGCGACCCGTA
     AGTACGCTCAGCTTCTCGTAGTGCTTCCCCCGTCCTGGCGGCCCGGGGCT
1200 GGGCTGCTCGCTGCTGCCGGTCACAGTCCCGCCAGCCGCGGAGCTGACTG
     AGCTCCCTTTCCCGGGACGTGTGCTCTGTGTTCGGTCAGCGAGGCTATCG
     GGAGGGCTTTGGCTGCATTTGGCTTCTCTGGCGCTTAGCGCAGGAGCACG
     TTGTGCTACGCCTGAACTACAGCTGTGAGAAGGCCGTGGAAACCGCTCTC
1400 AAACTGATTTATTGGCGAAATGGCTCTAAACTAAATCGTCTCCTCTCTTT
     GGAAATGCTTTAGAGAAGGTCTCTGTGGTAGTTCTTATGCATCTATCCTA
     AAGCACTTGGCCAGACAATTTAAAGACATCAAGCAGCATTTATAGCAGGC
     ACGTTTAATAACGAATACTGAATTTAAGTAACTCTGCTCACGTTGTATGA
1600 CGTTTATTTTCGTATTCCTGAAAGCCATTAAAATCCTGTGCAGTTGTTTA
     GTAAGAACAGCTGCCACTGTTTTGTATCTAGGAGATAACTGGTGTTTCCC
     TACAGTTCTCAAGCTGATAAAACTCTGTCTTTGTATCTAGGTAACCCTGT
     ATCACTTGCTGAAGCTTTTTCAGTCTGACACCAATGCAATCCTGGGAAAG
1800 AAAACTGTAGTTTCTGAATTCTATGATGAAATGGTATGAAAATTTTAATG
     TCAACCGAGCCTGACTTTATTTAAAAAAAATTATTGATGGTGCTGTGTAT
     TTTGGTCCTTCCTTAGATATTTCAAGATCCTACTGCCATGATGCAGCAAC
     TGCTAACGACGTCCCGTCAGCTGACACTTGGTGCTTACAAGCATGAAACA
```

FIG. 5A

```
2000  GAGTGTAAGTGCAAAATGAGGATACCTTCGCCGACCGTCATTCACTACTA
      ATGTTTTCTGTGGGATGTGATCGTACAGTGAGTTTGGCTGTGTGAAATTT
      GAATAGCTTGGTATTGGCAGTGATGACGTGATCGATGCCTTGCTTATCAT
      GTTTGAAATGAAGTAGAATAAATGCAGCCTGGTTTATTTGAGATAGTTTG
2200  GTTCATTTTATGGAATGCAAGCAAAGATTATACTTCCTCACTGAATTGCA
      CTGTCCAAAGGTGTGAAATGTGTGGGGATCTGGAGGACCGTGACCGAGGG
      ACATTGGATCGCTATCTCCCATTTCTTTTGCTGTTACCAGTTCAGATTTT
      CTTTTCACCTAGTCTTTAATTCCCAGGGTTTTGTTTTTTCCTTGGTCATA
2400  GTTTTTGTTTTTCACTCTGGCAAATGATGTTGTGAATTACACTGCTTCAG
      CCACAAAACTGATGGACTGAATGAGGTCATCAAACAAACTTTTCTTCTTC
      CGTATTTCCTTTTTTTTCCCCCACTTATCATTTTTACTGCTGTTGTTGAG
      TCTGTAAGGCTAAAAGTAACTGTTTTGTGCTTTTTCAGGACGTGTGCTTT
2600  CCAAATTACTGCCACATATATAAAGAAAGGTTGGAATTTTAAAGATAATT
      CATGTTTCTTCTTCTTTTTTGCCACCACAGTTGCAGATCTTGAAGTAAAA
      ACCAGGGAAAAGCTGGAAGCTGCCAAAAAGAAAACCAGTTTTGAAATTGC
      TGAGCTTAAAGAAAGGTTAAAAGCAAGTCGTGAAACCATCAACTGCTTAA
2800  AGAGTGAAATCAGAAAACTCGAAGAGGATGATCAGTCTAAAGATATGTGA
      TGAGTGTTGACTTGGCAGGGAGCCTATAATGAGAATGAAAGGACTTCAGT
      CGTGGAGTTGTATGCGTTCTCTCCAATTCTGTAACGGAGACTGTATGAAT
      TTCATTTGCAAATCACTGCAGTGTGTGACAACTGACTTTTTATAAATGGC
3000  AGAAAACAAGAATGAATGTATCCTCATTTTATAGTTAAAATCTATGGGTA
      TGTACTGGTTTATTTCAAGGAGAATGGATCGTAGAGACTTGGAGGCCAGA
      TTGCTGCTTGTATTGACTGCATTTGAGTGGTGTAGGAACATTTTGTCTAT
      GGTCCCGTGTTAGTTTACAGAATGCCACTGTTCACTGTTTTGTTTTGTAT
3200  TTTACTTTTTCTACTGCAACGTCAAGGTTTTAAAAGTTGAAAATAAAACA
      TGCAGGTTTTTTTTAAATATTTTTTTGTCTCTATCCAGTTTGGGCTTCAA
      GTATTATTGTTAACAGCAAGTCCTGATTTAAGTCAGAGGCTGAAGTGTAA
      TGGTATTCAAGATGCTTAAGTCTGTTGTCAGCAAAACAAAAGAGAAAACT
3400  TCATAAAATCAGGAAGTTGGCATTTCTAATAACTTCTTTATCAACAGATA
      AGAGTTTCTAGCCCTGCATCTACTTTCACTTATGTAGTTGATGCCTTTAT
      ATTTTGTGTGTTTGGATGCAGGAAGTGATTCCTACTCTGTTATGTAGATA
      TTCTATTTAACACTTGTACTCTGCTGTGCTTAGCCTTTCCCCATGAAAAT
3600  TCAGCGGCTGTAAATCCCCCTCTTCTTTTGTAGCCTCATACAGATGGCAG
      ACCCTCAGGCTTATAAAGGCTTGGGCATCTTCTTTACTGCTTTGAGATTC
      TGTGTTGCAGTAACCTCTGCCAGAGAGGAGAAAAGCCCCACAAACCTCAT
      CCCCTTCTTCTATAGCAATCAGTATTACTAATGCTTTGAGAACAGAGCAC
3800  TGGTTTGAAACGTTTGATAATTAGCATTTAACATGGCTTGGTAAAGATGC
      AGAACTGAAACAGCTGTGACAGTATGAACTCAGTATGGAGACTTCATTAA
      GACAAACAGCTGTTAAAATCAGGCATGTTTCATTGAGGAGGACGGGGCAA
      CTTGCACCAGTGGTGCCCACACAAATCCTTCCTGGCGCTGCAGACCAATT
```

FIG. 5B

```
4000  TTTCTGGCATTCTGACTGCCGTTGCTGCTGGTCACAGAGAGCAACTATTT
      TTATCAGCCACAGGCAATTTGCTTGTAGTATTTTCCAAGTGTTGTAGGTA
      AGTATAAATGCATCGGCTCCAGAGCACTTTGAGTATACTTATTAAAAACA
      TAAATGAAAGACAAATTAGCTTTGCTTGGGTGCACAGAACATTTTTAGTT
4200  CCAGCCTGCTTTTTGGTAGAAGCCCTCTTCTGAGGCTAGAACTGACTTTG
      ACAAGTAGAGAAACTGGCAACGGAGCTATTGCTATCGAAGGATCCTTGTT
      AACAAAGTTAATCGTCTTTTAAGGTTTGGTTTATTCATTAAATTTGCTTT
      TAAGCTGTAGCTGAAAAAGAACGTGCTGTCTTCCATGCACCAGGTGGCAG
4400  CTCTGTGCAAAGTGCTCTCTGGTCTCACCAGCCTTTTAATTGCCGGGATT
      CTGGCACGTCTGAGAGGGCTCAGACTGGCTTCGTTTGTTTGAACAGCGTG
      TACTGCTTTCTGTAGACATGGCCGGTTTCTCTCCTGCAGCTTATGAAACT
      GTTCACACTGAACACACTGGAACAGGTTGCCCAAGGAGGCCGTGGATGCC
4600  CCATCCCTGGAGGCATTCAAGGCCAGGCTGGATGTGGCTCTGGGCAGCCT
      GGTCTGGTGGTTGGCGATCCTGCACATAGCAGCGGGGTTGAAACTCGATG
      ATCACTGTGGTCCTTTTCAACCCAGGCTATTCTATGATTCTATGATTCAA
      CAGCAAATCATATGTACTGAGAGAGGAAACAAACACAAGTGCTACTGTTT
4800  GCAAGTTTTGTTCATTTGGTAAAAGAGTCAGGTTTTAAAATTCAAAATCT
      GTCTGGTTTTGGTGTTTTTTTTTTTTTATTTATTATTTCTTTGGGGTTCT
      TTTTGATGCTTTATCTTTCTCTGCCAGGACTGTGTGACAATGGGAACGAA
      AAAGAACATGCCAGGCACTGTCCTGGATTGCACACGCTGGTTGCACTCAG
5000  TAGCAGGCTCAGAACTGCCAGTCTTTCCACAGTATTACTTTCTAAACCTA
      ATTTTAATAGCGTTAGTAGACTTCCATCACTGGGCAGTGCTTAGTGAATG
      CTCTGTGTGAACGTTTTACTTATAAGCATGTTGGAAGTTTTGATGTTCCT
      GGATGCAGTAGGGAAGGACAGATTAGCTATGTGAAAAGTAGATTCTGAGT
5200  ATCGGGGTTACAAAAAGTATAGAAACGATGAGAAATTCTTGTTGTAACTA
      ATTGGAATTTCTTTAAGCGTTCACTTATGCTACATTCATAGTATTTCCAT
      TTAAAAGTAGGAAAAGGTAAAACGTGAAATCGTGTGATTTTCGGATGGAA
      CACCGCCTTCCTATGCACCTGACCAACTTCCAGAGGAAAAGCCTATTGAA
5400  AGCCGAGATTAAGCCACCAAAAGAACTCATTTGCATTGGAATATGTAGTA
      TTTGCCCTCTTCCTCCCGGGTAATTACTATACTTTATAGGGTGCTTATAT
      GTTAAATGAGTGGCTGGCACTTTTTATTCTCACAGCTGTGGGGAATTCTG
      TCCTCTAGGACAGAAACAATTTTAATCTGTTCCACTGGTGACTGCTTTGT
5600  CAGCACTTCCACCTGAAGAGATCAATACACTCTTCAATGTCTAGTTCTGC
      AACACTTGGCAAACCTCACATCTTATTTCATACTCTCTTCATGCCTATGC
      TTATTAAAGCAATAATCTGGGTAATTTTTGTTTTAATCACTGTCCTGACC
      CCAGTGATGACCGTGTCCCACCTAAAGCTCAATTCAGGTCCTGAATCTCT
5800  TCAACTCTCTATAGCTAACATGAAGAATCTTCAAAAGTTAGGTCTGAGGG
      ACTTAAGGCTAACTGTAGATGTTGTTGCCTGGTTTCTGTGCTGAAGGCCG
      TGTAGTAGTTAGAGCATTCAACCTCTAG
```

FIG. 5C

SEQ ID NO: 10

```
   1 TGCCGCCTTCTTTGATATTCACTCTGTTGTATTTCATCTCTTCTTGCCGA
     TGAAAGGATATAACAGTCTGTATAACAGTCTGTGAGGAAATACTTGGTAT
     TTCTTCTGATCAGTGTTTTTATAAGTAATGTTGAATATTGGATAAGGCTG
 151 TGTGTCCTTTGTCTTGGGAGACAAAGCCCACAGCAGGTGGTGGTTGGGGT
     GGTGGCAGCTCAGTGACAGGAGAGGTTTTTTTGCCTGTTTTTTTTTTTTT
     TTTTTTTTTTAAGTAAGGTGTTCTTTTTTCTTAGTAAATTTTCTACTGGA
 301 CTGTATGTTTTGACAGGTCAGAAACATTTCTTCAAAAGAAGAACCTTTTG
     GAAACTGTACAGCCCTTTTCTTTCATTCCCTTTTTGCTTTCTGTGCCAAT
     GCCTTTGGTTCTGATTGCATTATGGAAAACGTTGATCGGAACTTGAGGTT
 451 TTTATTTATAGTGTGGCTTGAAAGCTTGGATAGCTGTTGTTACACGAGAT
     ACCTTATTAAGTTTAGGCCAGCTTGATGCTTTATTTTTTCCCTTTGAAGT
     AGTGAGCGTTCTCTGGTTTTTTTCCTTTGAAACTGGTGAGGCTTAGATTT
 601 TTCTAATGGGATTTTTTACCTGATGATCTAGTTGCATACCCAAATGCTTG
     TAAATGTTTTCCTAGTTAACATGTTGATAACTTCGGATTTACATGTTGTA
     TATACTTGTCATCTGTGTTTCTAGTAAAAATATATGGCATTTATAGAAAT
 751 ACGTAATTCCTGATTTCCTTTTTTTTTATCTCTATGCTCTGTGTGTACAG
     GTCAAACAGACTTCACTCCTATTTTTATTTATAGAATTTTATATGCAGTC
     TGTCGTTGGTTCTTGTGTTGTAAGGATACAGCCTTAAATTTCCTAGAGCG
 901 ATGCTCAGTAAGGCGGGTTGTCACATGGGTTCAAATGTAAAACGGGCACG
     TTTGGCTGCTGCCTTCCCGAGATCCAGGACACTAAACTGCTTCTGCACTG
     AGGTATAAATCGCTTCAGATCCCAGGGAAGTGCAGATCCACGTGCATATT
1051 CTTAAAGAAGAATGAATACTTTCTAAAATATTTTGGCATAGGAAGCAAGC
     TGCATGGATTTGTTTGGGACTTAAATTATTTTGGTAACGGAGTGCATAGG
     TTTTAAACACAGTTGCAGCATGCTAACGAGTCACAGCGTTTATGCAGAAG
1201 TGATGCCTGGATGCCTGTTGCAGCTGTTTACGGCACTGCCTTGCAGTGAG
     CATTGCAGATAGGGGTGGGGTGCTTTGTGTCGTGTTCCCACACGCTGCCA
     CACAGCCACCTCCCGGAACACATCTCACCTGCTGGGTACTTTTCAAACCA
1351 TCTTAGCAGTAGTAGATGAGTTACTATGAAACAGAGAAGTTCCTCAGTTG
     GATATTCTCATGGGATGTCTTTTTTCCCATGTTGGGCAAAGTATGATAAA
     GCATCTCTATTTGTAAATTATGCACTTGTTAGTTCCTGAATCCTTTCTAT
1501 AGCACCACTTATTGCAGCAGGTGTAGGCTCTGGTGTGGCCTGTGTCTGTG
     CTTCAATCTTTTAAAGCTTCTTTGGAAATACACTGACTTGATTGAAGTCT
     CTTGAAGATAGTAAACAGTACTTACCTTTGATCCCAATGAAATCGAGCAT
1651 TTCAGTTGTAAAAGAATTCCGCCTATTCATACCATGTAATGTAATTTTAC
     ACCCCCAGTGCTGACACTTTGGAATATATTCAAGTAATAGACTTTGGCCT
     CACCCTCTTGTGTACTGTATTTTGTAATAGAAAATATTTTAAACTGTGCA
1801 TATGATTATTACATTATGAAAGAGACATTCTGCTGATCTTCAAATGTAAG
     AAAATGAGGAGTGCGTGTGCTTTTATAAATACAAGTGATTGCAAATTAGT
     GCAGGTGTCCTTAAAAAAAAAAAAAAAAAGTAATATAAAAAGGACCAGGT
1951 GTTTTACAAGTGAAATACATTCCTATTTGGTAAACAGTTACATTTTTATG
     AAGATTACCAGCGCTGCTGACTTTCTAAACATAAGGCTGTATTGTCTTCC
     TGTACCATTGCATTTCCTCATTCCCAATTTGCACAAGGATGTCTGGGTAA
2101 ACTATTCAAGAAATGGCTTTGAAATACAGCATGGGAGCTTGTCTGAGTTG
     GAATGCAGAGTTGCACTGCAAAATGTCAGGAAATGGATGTCTCTCAGAAT
     GCCCAACTCCAAAGGATTTTATATGTGTATATAGTAAGCAGTTTCCTGAT
```

FIG. 6A

```
2251  TCCAGCAGGCCAAAGAGTCTGCTGAATGTTGTGTTGCCGGAGACCTGTAT
      TTCTCAACAAGGTAAGATGGTATCCTAGCAACTGCGGATTTTAATACATT
      TTCAGCAGAAGTACTTAGTTAATCTCTACCTTTAGGGATCGTTTCATCAT
2401  TTTTAGATGTTATACTTGAAATACTGCATAACTTTTAGCTTTCATGGGTT
      CCTTTTTTTCAGCCTTTAGGAGACTGTTAAGCAATTTGCTGTCCAACTTT
      TGTGTTGGTCTTAAACTGCAATAGTAGTTTACCTTGTATTGAAGAAATAA
2551  AGACCATTTTTATATTAAAAAATACTTTTGTCTGTCTTCATTTTGACTTG
      TCTGATATCCTTGCAGTGCCCATTATGTCAGTTCTGTCAGATATTCAGAC
      ATCAAAACTTAACGTGAGCTCAGTGGAGTTACAGCTGCGGTTTTGATGCT
2701  GTTATTATTTCTGAAACTAGAAATGATGTTGTCTTCATCTGCTCATCAAA
      CACTTCATGCAGAGTGTAAGGCTAGTGAGAAATGCATACATTTATTGATA
      CTTTTTTAAAGTCAACTTTTTATCAGATTTTTTTTTCATTTGGAAATATA
2851  TTGTTTTCTAGACTGCATAGCTTCTGAATCTGAAATGCAGTCTGATTGGC
      ATGAAGAAGCACAGCACTCTTCATCTTACTTAAACTTCATTTTGGAATGA
      AGGAAGTTAAGCAAGGGCACAGGTCCATGAAATAGAGACAGTGCGCTCAG
3001  GAGAAAGTGAACCTGGATTTCTTTGGCTAGTGTTCTAAATCTGTAGTGAG
      GAAAGTAACACCCGATTCCTTGAAAGGGCTCCAGCTTTAATGCTTCCAAA
      TTGAAGGTGGCAGGCAACTTGGCCACTGGTTATTTACTGCATTATGTCTC
3151  AGTTTCGCAGCTAACCTGGCTTCTCCACTATTGAGCATGGACTATAGCCT
      GGCTTCAGAGGCCAGGTGAAGGTTGGGATGGGTGGAAGGAGTGCTGGGCT
      GTGGCTGGGGGGACTGTGGGGACTCCAAGCTGAGCTTGGGGTGGGCAGCA
3301  CAGGGAAAAGTGTGGGTAACTATTTTTAAGTACTGTGTTGCAAACGTCTC
      ATCTGCAAATACGTAGGGTGTGTACTCTCGAAGATTAACAGTGTGGGTTC
      AGTAATATATGGATGAATTCACAGTGGAAGCATTCAAGGGTAGATCATCT
3451  AACGACACCAGATCATCAAGCTATGATTGGAAGCGGTATCAGAAGAGCGA
      GGAAGGTAAGCAGTCTTCATATGTTTTCCCTCCACGTAAAGCAGTCTGGG
      AAAGTAGCACCCCTTGAGCAGAGACAAGGAAATAATTCAGGAGCATGTGC
3601  TAGGAGAACTTTCTTGCTGAATTCTACTTGCAAGAGCTTTGATGCCTGGC
      TTCTGGTGCCTTCTGCAGCACCTGCAAGGCCCAGAGCCTGTGGTGAGCTG
      GAGGGAAAGATTCTGCTCAAGTCCAAGCTTCAGCAGGTCATTGTCTTTGC
3751  TTCTTCCCCCAGCACTGTGCAGCAGAGTGGAACTGATGTCGAAGCCTCCT
      GTCCACTACCTGTTGCTGCAGGCAGACTGCTCTCAGAAAAAGAGAGCTAA
      CTCTATGCCATAGTCTGAAGGTAAAATGGGTTTTAAAAAAGAAAACACAA
3901  AGGCAAAACCGGCTGCCCCATGAGAAGAAAGCAGTGGTAAACATGGTAGA
      AAAGGTGCAGAAGCCCCCAGGCAGTGTGACAGGCCCCTCCTGCCACCTAG
      AGGCGGGAACAAGCTTCCCTGCCTAGGGCTCTGCCCGCGAAGTGCGTGTT
4051  TCTTTGGTGGGTTTTGTTTGGCGTTTGGTTTTGAGATTTAGACACAAGGG
      AAGCCTGAAAGGAGGTGTTGGGCACTATTTTGGTTTGTAAAGCCTGTACT
      TCAAATATATATTTTGTGAGGGAGTGTAGCGAATTGGCCAATTTAAAATA
4201  AAGTTGCAAGAGATTGAAGGCTGAGTAGTTGAGAGGGTAACACGTTTAAT
      GAGATCTTCTGAAACTACTGCTTCTAAACACTTGTTTGAGTGGTGAGACC
      TTGGATAGGTGAGTGCTCTTGTTACATGTCTGATGCACTTGCTTGTCCTT
4351  TTCCATCCACATCCATGCATTCCACATCCACGCATTTGTCACTTATCCCA
      TATCTGTCATATCTGACATACCTGTCTCTTCGTCACTTGGTCAGAAGAAA
      CAGATGTGATAATCCCCAGCCGCCCCAAGTTTGAGAAGATGGCAGTTGCT
4501  TCTTTCCCTTTTTCCTGCTAAGTAAGGATTTTCTCCTGGCTTTGACACCT
      CACGAAATAGTCTTCCTGCCTTACATTCTGGGCATTATTTCAAATATCTT
```

FIG. 6B

```
      TGGAGTGCGCTGCTCTCAAGTTTGTGTCTTCCTACTCTTAGAGTGAATGC
4651  TCTTAGAGTGAAAGAGAAGGAAGAGAAGATGTTGGCCGCAGTTCTCTGAT
      GAACACACCTCTGAATAATGGCCAAAGGTGGGTGGGTTTCTCTGAGGAAC
      GGGCAGCGTTTGCCTCTGAAAGCAAGGAGCTCTGCGGAGTTGCAGTTATT
4801  TTGCAACTGATGGTGGAACTGGTGCTTAAAGCAGATTCCCTAGGTTCCCT
      GCTACTTCTTTTCCTTCTTGGCAGTCAGTTTATTTCTGACAGACAAACAG
      CCACCCCCACTGCAGGCTTAGAAAGTATGTGGCTCTGCCTGGGTGTGTTA
4951  CAGCTCTGCCCTGGTGAAAGGGGATTAAAACGGGCACCATTCATCCCAAA
      CAGGATCCTCATTCATGGATCAAGCTGTAAGGAACTTGGGCTCCAACCTC
      AAAACATTAATTGGAGTACGAATGTAATTAAAACTGCATTCTCGCATTCC
5101  TAAGTCATTTAGTCTGGACTCTGCAGCATGTAGGTCGGCAGCTCCCACTT
      TCTCAAAGACCACTGATGGAGGAGTAGTAAAAATGGAGACCGATTCAGAA
      CAACCAACGGAGTGTTGCCGAAGAAACTGATGGAAATAATGCATGAATTG
5251  TGTGGTGGACATTTTTTTTAAATACATAAACTACTTCAAATGAGGTCGGA
      GAAGGTCAGTGTTTTATTAGCAGCCATAAAACCAGGTGAGCGAGTACCAT
      TTTTCTCTACAAGAAAAACGATTCTGAGCTCTGCGTAAGTATAAGTTCTC
5401  CATAGCGGCTGAAGCTCCCCCCTGGCTGCCTGCCATCTCAGCTGGAGTGC
      AGTGCCATTTCCTTGGGGTTTCTCTCACAGCAGTAATGGGACAATACTTC
      ACAAAAATTCTTTCTTTTCCTGTCATGTGGGATCCCTACTGTGCCCTCCT
5551  GGTTTTACGTTACCCCCTGACTGTTCCATTCAGCGGTTTGGAAAGAGAAA
      AAGAATTTGGAAATAAAACATGTCTACGTTATCACCTCCTCCAGCATTTT
      GGTTTTTAATTATGTCAATAACTGGCTTAGATTTGGAAATGAGAGGGGGT
5701  TGGGTGTATTACCGAGGAACAAAGGAAGGCTTATATAAACTCAAGTCTTT
      TATTTAGAGAACTGGCAAGCTGTCAAAAACAAAAAGGCCTTACCACCAAA
      TTAAGTGAATAGCCGCTATAGCCAGCAGGGCCAGCACGAGGGATGGTGCA
5851  CTGCTGGCACTATGCCACGGCCTGCTTGTGACTCTGAGAGCAACTGCTTT
      GGAAATGACAGCACTTGGTGCAATTTCCTTTGTTTCAGAATGCGTAGAGC
      GTGTGCTTGGCGACAGTTTTTCTAGTTAGGCCACTTCTTTTTTCCTTCTC
6001  TCCTCATTCTCCTAAGCATGTCTCCATGCTGGTAATCCCAGTCAAGTGAA
      CGTTCAAACAATGAATCCATCACTGTAGGATTCTCGTGGTGATCAAATCT
      TTGTGTGAGGTCTATAAAATATGGAAGCTTATTTATTTTTCGTTCTTCCA
6151  TATCAGTCTTCTCTATGACAATTCACATCCACCACAGCAAATTAAAGGTG
      AAGGAGGCTGGTGGGATGAAGAGGGTCTTCTAGCTTTACGTTCTTCCTTG
      CAAGGCCACAGGAAAATGCTGAGAGCTGTAGAATACAGCCTGGGGTAAGA
6301  AGTTCAGTCTCCTGCTGGGACAGCTAACCGCATCTTATAACCCCTTCTGA
      GACTCATCTTAGGACCAAATAGGGTCTATCTGGGGTTTTTGTTCCTGCTG
      TTCCTCCTGGAAGGCTATCTCACTATTTCACTGCTCCCACGGTTACAAAC
6451  CAAAGATACAGCCTGAATTTTTTCTAGGCCACATTACATAAATTTGACCT
      GGTACCAATATTGTTCTCTATATAGTTATTTCCTTCCCCACTGTGTTTAA
      CCCCTTAAGGCATTCAGAACAACTAGAATCATAGAATGGTTTGGATTGGA
6601  AGGGGCCTTAAACATCATCCATTTCCAACCCTCTGCCATGGGCTGCTTGC
      CACCCACTGGCTCAGGCTGCCCAGGGCCCCATCCAGCCTGGCCTTGAGCA
      CCTCCAGGGATGGGGCACCCACAGCTTCTCTGGGCAGCCTGTGCCAACAC
6751  CTCACCACTCTCTGGGTAAAGAATTCTCTTTTAACATCTAATCTAAATCT
      CTTCTCTTTTAGTTTAAAGCCATTCCTCTTTTTCCCGTTGCTATCTGTCC
      AAGAAATGTGTATTGGTCTCCCTCCTGCTTATAAGCAGGAAGTACTGGAA
6901  GGCTGCAGTGAGGTCTCCCCACAGCCTTCTCTTCTCCAGGCTGAACAAGC
```

FIG. 6C

```
      CCAGCTCCTTCAGCCTGTCTTCGTAGGAGATCATCTTAGTGGCCCTCCTC
      TGGACCCATTCCAACAGTTCCACGGCTTTCTTGTGGAGCCCCAGGTCTGG
7051  ATGCAGTACTTCAGATGGGGCCTTACAAAGGCAGAGCAGATGGGGACAAT
      CGCTTACCCCTCCCTGCTGGCTGCCCCTGTTTTGATGCAGCCCAGGGTAC
      TGTTGGCCTTTCAGGCTCCCAGACCCCTTGCTGATTTGTGTCAAGCTTTT
7201  CATCCACCAGAACCCACGCTTCCTGGTTAATACTTCTGCCCTCACTTCTG
      TAAGCTTGTTTCAGGAGACTTCCATTCTTTAGGACAGACTGTGTTACACC
      TACCTGCCCTATTCTTGCATATATACATTTCAGTTCATGTTTCCTGTAAC
7351  AGGACAGAATATGTATTCCTCTAACAAAAATACATGCAGAATTCCTAGTG
      CCATCTCAGTAGGGTTTTCATGGCAGTATTAGCACATAGTCAATTTGCTG
      CAAGTACCTTCCAAGCTGCGGCCTCCCATAAATCCTGTATTTGGGATCAG
7501  TTACCTTTTGGGGTAAGCTTTTGTATCTGCAGAGACCCTGGGGGGTTCTGA
      TGTGCTTCAGCTCTGCTCTGTTCTGACTGCACCATTTTCTAGATCACCCA
      GTTGTTCCTGTACAACTTCCTTGTCCTCCATCCTTTCCCAGCTTGTATCT
7651  TTGACAAATACAGGCCTATTTTTGTGTTTGCTTCAGCAGCCATTTAATTC
      TTCAGTGTCATCTTGTTCTGTTGATGCCACTGGAACAGGATTTTCAGCAG
      TCTTGCAAAGAACATCTAGCTGAAAACTTTCTGCCATTCAATATTCTTAC
7801  CAGTTCTTCTTGTTTGAGGTGAGCCATAAATTACTAGAACTTCGTCACTG
      ACAAGTTTATGCATTTTATTACTTCTATTATGTACTTACTTTGACATAAC
      ACAGACACGCACATATTTTGCTGGGATTTCCACAGTGTCTCTGTGTCCTT
7951  CACATGGTTTTACTGTCATACTTCCGTTATAACCTTGGCAATCTGCCCAG
      CTGCCCATCACAAGAAAAGAGATTCCTTTTTTATTACTTCTCTTCAGCCA
      ATAAACAAAATGTGAGAAGCCCAAACAAGAACTTGTGGGGCAGGCTGCCA
8101  TCAAGGGAGAGACAGCTGAAGGGTTGTGTAGCTCAATAGAATTAAGAAAT
      AATAAAGCTGTGTCAGACAGTTTTGCCTGATTTATACAGGCACGCCCCAA
      GCCAGAGAGGCTGTCTGCCAAGGCCACCTTGCAGTCCTTGGTTTGTAAGA
8251  TAAGTCATAGGTAACTTTTCTGGTGAATTGCGTGGAGAATCATGATGGCA
      GTTCTTGCTGTTTACTATGGTAAGATGCTAAAATAGGAGACAGCAAAGTA
      ACACTTGCTGCTGTAGGTGCTCTGCTATCCAGACAGCGATGGCACTCGCA
8401  CACCAAGATGAGGGATGCTCCCAGCTGACGGATGCTGGGGCAGTAACAGT
      GGGTCCCATGCTGCCTGCTCATTAGCATCACCTCAGCCCTCACCAGCCCA
      TCAGAAGGATCATCCCAAGCTGAGGAAAGTTGCTCATCTTCTTCACATCA
8551  TCAAACCTTTGGCCTGACTGATGGCTCCCGGATGCTTAAATGTGGTCACT
      GACATCTTTATTTTTCTATGATTTCAAGTCAGAACCTCCGGATCAGGAGG
      GAACACATAGTGGGAATGTACCCTCAGCTCCAAGGCCAGATCTTCCTTCA
8701  ATGATCATGCATGCTACTTAGGAAGGTGTGTGTGTGTGAATGTAGAATTG
      CCTTTGTTATTTTTTCTTCCTGCTGTCAGGAACATTTTGAATACCAGAGA
      AAAAGAAAAGTGCTCTTCTTGGCATGGGAGGAGTTGTCACACTTGCAAAA
8851  TAAAGGATGCAGTCCCAAATGTTCATAATCTCAGGGTCTGAAGGAGGATC
      AGAAACTGTGTATACAATTTCAGGCTTCTCTGAATGCAGCTTTTGAAAGC
      TGTTCCTGGCCGAGGCAGTACTAGTCAGAACCCTCGGAAACAGGAACAAA
9001  TGTCTTCAAGGTGCAGCAGGAGGAAACACCTTGCCCATCATGAAAGTGAA
      TAACCACTGCCGCTGAAGGAATCCAGCTCCTGTTTGAGCAGGTGCTGCAC
      ACTCCCACACTGAAACAACAGTTCATTTTTATAGGACTTCCAGGAAGGAT
9151  CTTCTTCTTAAGCTTCTTAATTATGGTACATCTCCAGTTGGCAGATGACT
      ATGACTACTGACAGGAGAATGAGGAACTAGCTGGGAATATTTCTGTTTGA
      CCACCATGGAGTCACCCATTTCTTTACTGGTATTTGGAAATAATAATTCT
```

FIG. 6D

```
9301  GAATTGCAAAGCAGGAGTTAGCGAAGATCTTCATTTCTTCCATGTTGGTG
      ACAGCACAGTTCTGGCTATGAAAGTCTGCTTACAAGGAAGAGGATAAAAA
9401  TCATAGGGATAATAAATCTAAGTTTGAAGACAATGAGGTTTTAGCTGCAT
      TTGACATGAAGAAATTGAGACCTCTACTGGATAGCTATGGTATTTACGTG
      TCTTTTTGCTTAGTTACTTATTGACCCCAGCTGAGGTCAAGTATGAACTC
9551  AGGTCTCTCGGGCTACTGGCATGGATTGATTACATACAACTGTAATTTTA
      GCAGTGATTTAGGGTTTATGAGTACTTTTGCAGTAAATCATAGGGTTAGT
      AATGTTAATCTCAGGGAAAAAAAAAAAAAGCCAACCCTGACAGACATCCC
9701  AGCTCAGGTGGAAATCAAGGATCACAGCTCAGTGCGGTCCCAGAGAACAC
      AGGGACTCTTCTCTTAGGACCTTTATGTACAGGGCCTCAAGATAACTGAT
      GTTAGTCAGAAGACTTTCCATTCTGGCCACAGTTCAGCTGAGGCAATCCT
9851  GGAATTTTCTCTCCGCTGCACAGTTCCAGTCATCCCAGTTTGTACAGTTC
      TGGCACTTTTTGGGTCAGGCCGTGATCCAAGGAGCAGAAGTTCCAGCTAT
      GGTCAGGGAGTGCCTGACCGTCCCAACTCACTGCACTCAAACAAAGGCGA
10001 AACCACAAGAGTGGCTTTTGTTGAAATTGCAGTGTGGCCCAGAGGGGCTG
      CACCAGTACTGGATTGACCACGAGGCAACATTAATCCTCAGCAAGTGCAA
      TTTGCAGCCATTAAATTGAACTAACTGATACTACAATGCAATCAGTATCA
10151 ACAAGTGGTTTGGCTTGGAAGATGGAGTCTAGGGGCTCTACAGGAGTAGC
      TACTCTCTAATGGAGTTGCATTTTGAAGCAGGACACTGTGAAAAGCTGGC
      CTCCTAAAGAGGCTGCTAAACATTAGGGTCAATTTTCCAGTGCACTTTCT
10301 GAAGTGTCTGCAGTTCCCCATGCAAAGCTGCCCAAACATAGCACTTCCAA
      TTGAATACAATTATATGCAGGCGTACTGCTTCTTGCCAGCACTGTCCTTC
      TCAAATGAACTCAACAAACAATTTCAAAGTCTAGTAGAAAGTAACAAGCT
10451 TTGAATGTCATTAAAAAGTATATCTGCTTTCAGTAGTTCAGCTTATTTAT
      GCCCACTAGAAACATCTTGTACAAGCTGAACACTGGGGCTCCAGATTAGT
      GGTAAAACCTACTTTATACAATCATAGAATCATAGAATGGCCTGGGTTGG
10601 AAGGGACCCCAAGGATCATGAAGATCCAACACCCCCGCCACAGGCAGGGC
      CACCAACCTCCAGATCTGGTACTAGACCAGGCAGCCCAGGGCTCCATCCA
      ACCTGGCCATGAACACCTCCAGGGATGGAGCATCCACAACCTCTCTGGGC
10751 AGCCTGTGCCAGCACCTCACCACCCTCTCTGTGAAGAACTTTTCCCTGAC
      ATCCAATCTAAGCCTTCCCTCCTTGAGGTTAGATCCACTCCCCCTTGTGC
      TATCACTGTCTACTCTTGTAAAAAGTTGATTCTCCTCCTTTTTGGAAGGT
10901 TGCAATGAGGTCTCCTTGCAGCCTTCTTCTCTTCTGCAGGATGAACAAGC
      CCAGCTCCCTCAGCCTGTCTTTATAGGAGAGGTGCTCCAGCCCTCTGATC
      ATCTTTGTGGCCCTCCTCTGGACCCGCTCCAAGAGCTCCACATCTTTCCT
11051 GTACTGGGGGCCCCAGGCCTGAATGCAGTACTCCAGATGGGGCCTCAAAA
      GAGCAGAGTAAAGAGGGACAATCACCTTCCTCACCCTGCTGGCCAGCCCT
      CTTCTGATGGAGCCCTGGATACAACTGGCTTTCTGAGCTGCAACTTCTCC
11201 TTATCAGTTCCACTATTAAAACAGGAACAATACAACAGGTGCTGATGGCC
      AGTGCAGAGTTTTTCACACTTCTTCATTTCGGTAGATCTTAGATGAGGAA
      CGTTGAAGTTGTGCTTCTGCGTGTGCTTCTTCCTCCTCAAATACTCCTGC
11351 CTGATACCTCACCCCACCTGCCACTGAATGGCTCCATGGCCCCCTGCAGC
      CAGGGCCCTGATGAACCCGGCACTGCTTCAGATGCTGTTTAATAGCACAG
      TATGACCAAGTTGCACCTATGAATACACAAACAATGTGTTGCATCCTTCA
11501 GCACTTGAGAAGAAGAGCCAAATTTGCATTGTCAGGAAATGGTTTAGTAA
      TTCTGCCAATTAAAACTTGTTTATCTACCATGGCTGTTTTTATGGCTGTT
      AGTAGTGGTACACTGATGATGAACAATGGCTATGCAGTAAAATCAAGACT
```

FIG. 6E

```
11651 GTAGATATTGCAACAGACTATAAAATTCCTCTGTGGCTTAGCCAATGTGG
      TACTTCCCACATTGTATAAGAAATTTGGCAAGTTTAGAGCAATGTTTGAA
      GTGTTGGGAAATTTCTGTATACTCAAGAGGGCGTTTTTGACAACTGTAGA
11801 ACAGAGGAATCAAAAGGGGGTGGGAGGAAGTTAAAAGAAGAGGCAGGTGC
      AAGAGAGCTTGCAGTCCCGCTGTGTGTACGACACTGGCAACATGAGGTCT
      TTGCTAATCTTGGTGCTTTGCTTCCTGCCCCTGGCTGCCTTAGGGTGCGA
11951 TCTGCCTCAGACCCACAGCCTGGGCAGCAGGAGGACCCTGATGCTGCTGG
      CTCAGATGAGGAGAATCAGCCTGTTTAGCTGCCTGAAGGATAGGCACGAT
      TTTGGCTTTCCTCAAGAGGAGTTTGGCAACCAGTTTCAGAAGGCTGAGAC
12101 CATCCCTGTGCTGCACGAGATGATCCAGCAGATCTTTAACCTGTTTAGCA
      CCAAGGATAGCAGCGCTGCTTGGGATGAGACCCTGCTGGATAAGTTTTAC
      ACCGAGCTGTACCAGCAGCTGAACGATCTGGAGGCTTGCGTGATCCAGGG
12251 CGTGGGCGTGACCGAGACCCCTCTGATGAAGGAGGATAGCATCCTGGCTG
      TGAGGAAGTACTTTCAGAGGATCACCCTGTACCTGAAGGAGAAGAAGTAC
      AGCCCCTGCGCTTGGGAAGTCGTGAGGGCTGAGATCATGAGGAGCTTTAG
12401 CCTGAGCACCAACCTGCAAGAGAGCTTGAGGTCTAAGGAGTAAAAAGTCT
      AGAGTCGGGGCGGCGCGTGGTAGGTGGCGGGGGGTTCCCAGGAGAGCCCC
      CAGCGCGGACGGCAGCGCCGTCACTCACCGCTCCGTCTCCCTCCGCCCAG
12551 GGTCGCCTGGCGCAACCGCTGCAAGGGCACCGACGTCCAGGCGTGGATCA
      GAGGCTGCCGGCTGTGAGGAGCTGCCGCGCCCGGCCCGCCCGCTGCACAG
      CCGGCCGCTTTGCGAGCGCGACGCTACCCGCTTGGCAGTTTTAAACGCAT
12701 CCCTCATTAAAACGACTATACGCAAACGCCTTCCCGTCGGTCCGCGTCTC
      TTTCCGCCGCCAGGGCGACACTCGCGGGGAGGGCGGGAAGGGGGCCGGGC
      GGGAGCCCGCGGCCAACCGTCGCCCCGTGACGGCACCGCCCCGCCCCCGT
12851 GACGCGGTGCGGGCGCCGGGGCCGTGGGGCTGAGCGCTGCGGCGGGGCCG
      GGCCGGGCCGGGGCGGGAGCTGAGCGCGGCGCGGCTGCGGGCGGCGCCCC
      CTCCGGTGCAATATGTTCAAGAGAATGGCTGAGTTCGGGCCTGACTCCGG
13001 GGGCAGGGTGAAGGTGCGGCGCGGGCGGAGGGACGGGGCGGGCGCGGGGC
      CGCCCGGCGGGTGCCGGGGCCTCTGCCGGCCCGCCCGGCTCGGGCTGCTG
      CGGCGCTTACGGGCGCGCTTCTCGCCGCTGCCGCTTCTCTTCTCTCCCGC
13151 GCAAGGGCGTCACCATCGTGAAGCCGGTAGTGTACGGGAACGTGGCGCGG
      TACTTCGGGAAGAAGAGGGAGGAGGACGGGCACACGCATCAGTGGACGGT
      TTACGTGAAGCCCTACAGGAACGAGGTAGGGCCCGAGCGCGTCGGCCGCC
13301 GTTCTCGGAGCGCCGGAGCCGTCAGCGCCGCGCCTGGGTGCGCTGTGGGA
      CACAGCGAGCTTCTCTCGTAGGACATGTCCGCCTACGTGAAAAAAATCCA
      GTTCAAGCTGCACGAGAGCTACGGGAATCCTCTCCGAGGTGGGTGTTGCG
13451 TCGGGGGGTTTGCTCCGCTCGGTCCCGCTGAGGCTCGTCGCCCTCATCTT
      TCTTTCGTGCCGCAGTCGTTACCAAACCGCCGTACGAGATCACCGAAACG
13551 GGCTGGGGCGAATTTGAAATCATCATCAAGATATTTTTCATTGATCCAAA
      CGAGCGACCCGTAAGTACGCTCAGCTTCTCGTAGTGCTTCCCCCGTCCTG
      GCGGCCCGGGGCTGGGCTGCTCGCTGCTGCCGGTCACAGTCCCGCCAGCC
13701 GCGGAGCTGACTGAGCTCCCTTTCCCGGGACGTGTGCTCTGTGTTCGGTC
      AGCGAGGCTATCGGGAGGGCTTTGGCTGCATTTGGCTTCTCTGGCGCTTA
      GCGCAGGAGCACGTTGTGCTACGCCTGAACTACAGCTGTGAGAAGGCCGT
13851 GGAAACCGCTCTCAAACTGATTTATTGGCGAAATGGCTCTAAACTAAATC
      GTCTCCTCTCTTTGGAAATGCTTTAGAGAAGGTCTCTGTGGTAGTTCTTA
      TGCATCTATCCTAAAGCACTTGGCCAGACAATTTAAAGACATCAAGCAGC
```

FIG. 6F

```
14001 ATTTATAGCAGGCACGTTTAATAACGAATACTGAATTTAAGTAACTCTGC
      TCACGTTGTATGACGTTTATTTTCGTATTCCTGAAAGCCATTAAAATCCT
      GTGCAGTTGTTTAGTAAGAACAGCTGCCACTGTTTTGTATCTAGGAGATA
14151 ACTGGTGTTTCCCTACAGTTCTCAAGCTGATAAAACTCTGTCTTTGTATC
      TAGGTAACCCTGTATCACTTGCTGAAGCTTTTTCAGTCTGACACCAATGC
      AATCCTGGGAAAGAAAACTGTAGTTTCTGAATTCTATGATGAAATGGTAT
14301 GAAAATTTTAATGTCAACCGAGCCTGACTTTATTTAAAAAAAATTATTGA
      TGGTGCTGTGTATTTTGGTCCTTCCTTAGATATTTCAAGATCCTACTGCC
      ATGATGCAGCAACTGCTAACGACGTCCCGTCAGCTGACACTTGGTGCTTA
14451 CAAGCATGAAACAGAGTGTAAGTGCAAAATGAGGATACCTTCGCCGACCG
      TCATTCACTACTAATGTTTTCTGTGGGATGTGATCGTACAGTGAGTTTGG
      CTGTGTGAAATTTGAATAGCTTGGTATTGGCAGTGATGACGTGATCGATG
14601 CCTTGCTTATCATGTTTGAAATGAAGTAGAATAAATGCAGCCTGCTTTAT
      TTGAGATAGTTTGGTTCATTTTATGGAATGCAAGCAAAGATTATACTTCC
      TCACTGAATTGCACTGTCCAAAGGTGTGAAATGTGTGGGGATCTGGAGGA
14751 CCGTGACCGAGGGACATTGGATCGCTATCTCCCATTTCTTTTGCTGTTAC
      CAGTTCAGATTTTCTTTTCACCTAGTCTTTAATTCCCAGGGTTTTGTTTT
      TTCCTTGGTCATAGTTTTTGTTTTTCACTCTGGCAAATGATGTTGTGAAT
14901 TACACTGCTTCAGCCACAAAACTGATGGACTGAATGAGGTCATCAAACAA
      ACTTTTCTTCTTCCGTATTTCCTTTTTTTTCCCCCACTTATCATTTTTAC
      TGCTGTTGTTGAGTCTGTAAGGCTAAAAGTAACTGTTTTGTGCTTTTTCA
15051 GGACGTGTGCTTTCCAAATTACTGCCACATATATAAAGAAAGGTTGGAAT
      TTTAAAGATAATTCATGTTTCTTCTTCTTTTTTGCCACCACAGTTGCAGA
      TCTTGAAGTAAAAACCAGGGAAAAGCTGGAAGCTGCCAAAAAGAAAACCA
15201 GTTTTGAAATTGCTGAGCTTAAAGAAAGGTTAAAAGCAAGTCGTGAAACC
      ATCAACTGCTTAAAGAGTGAAATCAGAAAACTCGAAGAGGATGATCAGTC
      TAAAGATATGTGATGAGTGTTGACTTGGCAGGGAGCCTATAATGAGAATG
15351 AAAGGACTTCAGTCGTGGAGTTGTATGCGTTCTCTCCAATTCTGTAACGG
      AGACTGTATGAATTTCATTTGCAAATCACTGCAGTGTGTGACAACTGACT
      TTTTATAAATGGCAGAAAACAAGAATGAATGTATCCTCATTTTATAGTTA
15501 AAATCTATGGGTATGTACTGGTTTATTTCAAGGAGAATGGATCGTAGAGA
      CTTGGAGGCCAGATTGCTGCTTGTATTGACTGCATTTGAGTGGTGTAGGA
      ACATTTTGTCTATGGTCCCGTGTTAGTTTACAGAATGCCACTGTTCACTG
15651 TTTTGTTTTGTATTTTACTTTTTCTACTGCAACGTCAAGGTTTTAAAAGT
      TGAAAATAAAACATGCAGGTTTTTTTTAAATATTTTTTTGTCTCTATCCA
      GTTTGGGCTTCAAGTATTATTGTTAACAGCAAGTCCTGATTTAAGTCAGA
15801 GGCTGAAGTGTAATGGTATTCAAGATGCTTAAGTCTGTTGTCAGCAAAAC
      AAAAGAGAAAACTTCATAAAATCAGGAAGTTGGCATTTCTAATAACTTCT
      TTATCAACAGATAAGAGTTTCTAGCCCTGCATCTACTTTCACTTATGTAG
15951 TTGATGCCTTTATATTTTGTGTGTTTGGATGCAGGAAGTGATTCCTACTC
      TGTTATGTAGATATTCTATTTAACACTTGTACTCTGCTGTGCTTAGCCTT
16051 TCCCCATGAAAATTCAGCGGCTGTAAATCCCCCTCTTCTTTTGTAGCCTC
      ATACAGATGGCAGACCCTCAGGCTTATAAAGGCTTGGGCATCTTCTTTAC
      TGCTTTGAGATTCTGTGTTGCAGTAACCTCTGCCAGAGAGGAGAAAAGCC
16201 CCACAAACCTCATCCCCTTCTTCTATAGCAATCAGTATTACTAATGCTTT
      GAGAACAGAGCACTGGTTTGAAACGTTTGATAATTAGCATTTAACATGGC
      TTGGTAAAGATGCAGAACTGAAACAGCTGTGACAGTATGAACTCAGTATG
```

FIG. 6G

```
16351 GAGACTTCATTAAGACAAACAGCTGTTAAAATCAGGCATGTTTCATTGAG
      GAGGACGGGGCAACTTGCACCAGTGGTGCCCACACAAATCCTTCCTGGCG
      CTGCAGACCAATTTTTCTGGCATTCTGACTGCCGTTGCTGCTGGTCACAG
16501 AGAGCAACTATTTTTATCAGCCACAGGCAATTTGCTTGTAGTATTTTCCA
      AGTGTTGTAGGTAAGTATAAATGCATCGGCTCCAGAGCACTTTGAGTATA
      CTTATTAAAAACATAAATGAAAGACAAATTAGCTTTGCTTGGGTGCACAG
16651 AACATTTTTAGTTCCAGCCTGCTTTTTGGTAGAAGCCCTCTTCTGAGGCT
      AGAACTGACTTTGACAAGTAGAGAAACTGGCAACGGAGCTATTGCTATCG
16751 AAGGATCCTTGTTAACAAAGTTAATCGTCTTTTAAGGTTTGGTTTATTCA
      TTAAATTTGCTTTTAAGCTGTAGCTGAAAAAGAACGTGCTGTCTTCCATG
16851 CACCAGGTGGCAGCTCTGTGCAAAGTGCTCTCTGGTCTCACCAGCCTTTT
      AATTGCCGGGATTCTGGCACGTCTGAGAGGGCTCAGACTGGCTTCGTTTG
      TTTGAACAGCGTGTACTGCTTTCTGTAGACATGGCCGGTTTCTCTCCTGC
17001 AGCTTATGAAACTGTTCACACTGAACACACTGGAACAGGTTGCCCAAGGA
      GGCCGTGGATGCCCCATCCCTGGAGGCATTCAAGGCCAGGCTGGATGTGG
      CTCTGGGCAGCCTGGTCTGGTGGTTGGCGATCCTGCACATAGCAGCGGGG
17151 TTGAAACTCGATGATCACTGTGGTCCTTTTCAACCCAGGCTATTCTATGA
      TTCTATGATTCAACAGCAAATCATATGTACTGAGAGAGGAAACAAACACA
      AGTGCTACTGTTTGCAAGTTTTGTTCATTTGGTAAAAGAGTCAGGTTTTA
17301 AAATTCAAAATCTGTCTGGTTTTGGTGTTTTTTTTTTTTTATTTATTATT
      TCTTTGGGGTTCTTTTTGATGCTTTATCTTTCTCTGCCAGGACTGTGTGA
      CAATGGGAACGAAAAAGAACATGCCAGGCACTGTCCTGGATTGCACACGC
17451 TGGTTGCACTCAGTAGCAGGCTCAGAACTGCCAGTCTTTCCACAGTATTA
      CTTTCTAAACCTAATTTTAATAGCGTTAGTAGACTTCCATCACTGGGCAG
      TGCTTAGTGAATGCTCTGTGTGAACGTTTTACTTATAAGCATGTTGGAAG
17601 TTTTGATGTTCCTGGATGCAGTAGGGAAGGACAGATTAGCTATGTGAAAA
      GTAGATTCTGAGTATCGGGGTTACAAAAAGTATAGAAACGATGAGAAATT
      CTTGTTGTAACTAATTGGAATTTCTTTAAGCGTTCACTTATGCTACATTC
17751 ATAGTATTTCCATTTAAAAGTAGGAAAAGGTAAAACGTGAAATCGTGTGA
      TTTTCGGATGGAACACCGCCTTCCTATGCACCTGACCAACTTCCAGAGGA
      AAAGCCTATTGAAAGCCGAGATTAAGCCACCAAAAGAACTCATTTGCATT
17901 GGAATATGTAGTATTTGCCCTCTTCCTCCCGGGTAATTACTATACTTTAT
      AGGGTGCTTATATGTTAAATGAGTGGCTGGCACTTTTTATTCTCACAGCT
      GTGGGGAATTCTGTCCTCTAGGACAGAAACAATTTTAATCTGTTCCACTG
18051 GTGACTGCTTTGTCAGCACTTCCACCTGAAGAGATCAATACACTCTTCAA
      TGTCTAGTTCTGCAACACTTGGCAAACCTCACATCTTATTTCATACTCTC
      TTCATGCCTATGCTTATTAAAGCAATAATCTGGGTAATTTTTGTTTTAAT
18201 CACTGTCCTGACCCCAGTGATGACCGTGTCCCACCTAAAGCTCAATTCAG
      GTCCTGAATCTCTTCAACTCTCTATAGCTAACATGAAGAATCTTCAAAAG
      TTAGGTCTGAGGGACTTAAGGCTAACTGTAGATGTTGTTGCCTGGTTTCT
18351 GTGCTGAAGGCCGTGTAGTAGTTAGAGCATTCAACCTCTAGAAGAAGCTT
      GGCCAGCTGGTCGACCTGCAGATCCGGCCCTCGAGGGGGGGCCCGGTACC
      CAGCTTTTGTTCCCTTTAGTGAGGGTTAATTTCGAGCTTGGCGTAATCAT
18501 GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACAC
      AACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGT
      GAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGG
18651 GAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGA
```

FIG. 6H

```
      GGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCT
      GCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG
18801 TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGA
      GCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGC
      GTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCT
18951 CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
      CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTAC
      CGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATA
19101 GCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTG
      GGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGG
      TAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG
19251 CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCT
      ACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGT
      ATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTG
19401 GTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTT
      GTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCC
      TTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTT
19551 AAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTT
      TTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAAC
      TTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGA
19701 TCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATA
      ACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACC
      GCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAG
19851 CCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATC
      CAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAA
      TAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCT
20001 CGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGA
      GTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCC
      TCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTA
20151 TGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTT
      TCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG
      GCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCAC
20301 ATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGA
      AAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCAC
      TCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTG
20451 GGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCG
      ACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAG
      CATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT
20601 AGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCA
      CCTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGT
      TAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTA
20751 TAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGA
      ACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAA
      ACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAG
20901 TTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGA
      GCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAG
      GAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGC
```

FIG. 6I

```
21051 GGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTAC
      AGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGA
      TCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGC
21201 TGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTT
      GTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTG
21301 GAGCTCCACCGCGGTGGCGGCCGCTCTAG
```

FIG. 6J

SEQ ID NO: 11

GTACCGGGCCCCCCCTCGAGGTGAATATCCAAGAATGCAGAACTGCATGGAAAGCAGAGCTG
CAGGCACGATGGTGCTGAGCCTTAGCTGCTTCCTGCTGGGAGATGTGGATGCAGAGACGAAT
GAAGGACCTGTCCCTTACTCCCCTCAGCATTCTGTGCTATTTAGGGTTCTACCAGAGTCCTT
AAGAGGTTTTTTTTTTTTTGGTCCAAAAGTCTGTTTGTTTGGTTTTGACCACTGAGAGCAT
GTGACACTTGTCTCAAGCTATTAACCAAGTGTCCAGCCAAAATCGATGTCACAACTTGGGAA
TTTTCCATTTGAAGCCCCTTGCAAAAACAAAGAGCACCTTGCCTGCTCCAGCTCCTGGCTGT
GAAGGGTTTTGGTGCCAAAGAGTGAAAGGCTTCCTAAAAATGGGCTGAGCCGGGGAAGGGGG
GCAACTTGGGGGCTATTGAGAAACAAGGAAGGACAAACAGCGTTAGGTCATTGCTTCTGCAA
ACACAGCCAGGGCTGCTCCTCTATAAAAGGGGAAGAAAGAGGCTCCGCAGCCATCACAGACC
CAGAGGGGACGGTCTGTGAATCAAGCTT

FIG. 7

SEQ ID NO. 14
IFN-A
ATGGCTTTGA CCTTTGCCTT ACTGGTGGCT CTCCTGGTGC TGAGCTGCAA GAGCAGCTGC
TCTGTGGGCT GCGATCTGCC TCA

SEQ ID NO. 15
IFN-B
GACCCACAGC CTGGGCAGCA GGAGGACCCT GATGCTGCTG GCTCAGATGA GGAGAATCAG
CCTGTTTAGC TGCCTGAAGG ATAGGCACGA TTTTGGCTTT

SEQ ID NO. 16
IFN-C
CTCAAGAGGA GTTTGGCAAC CAGTTTCAGA AGGCTGAGAC CATCCCTGTG CTGCACGAGA
TG

SEQ ID NO. 17
IFN-D
TCCAGCAGAT CTTTAACCTG TTTAGCACCA AGGATAGCAG CGCTGCTTGG GATGAGACCC
TGCTGGATAA GTTTTACACC GAGCTGTACC AGCA

SEQ ID NO. 18
IFN-E
CTGAACGATC TGGAGGCTTG CGTGATCCAG GGCGTGGGCG TGACCGAGAC CCCTCTGATG
AAGGAGGATA GCATCCT

SEQ ID NO. 19
UFN-F
GCTGTGAGGA AGTACTTTCA GAGGATCACC CTGTACCTGA AGGAGAAGAA GTACAGCCCT
TGCGCTTGGG AAGTCGTGAG GG

SEQ ID NO. 20
IFN-G
CTGAGATCAT GAGGAGCTTT AGCCTGAGCA CCAACCTGCA AGAGAGCTTG AGGTCTAAGG
AGTAA

SEQ ID NO. 21
IFN-1
CCCAAGCTTT CACCATGGCT TTGACCTTTG CCTT

SEQ ID NO. 22
IFN-2b
ATCTGCCTCA GACCCACAG

FIG. 8A

SEQ ID NO. 23
IFN-3c
GATTTTGGCT TTCCTCAAGA GGAGTT

SEQ ID NO. 24
IFN-4b
GCACGAGATG ATCCAGCAGA T

SEQ ID NO. 25
IFN-5
ATCGTTCAGC TGCTGGTACA

SEQ ID NO. 26
IFN-6
CCTCACAGCC AGGATGCTAT

SEQ ID NO. 27
IFN-7
ATGATCTCAG CCCTCACGAC

SEQ ID NO. 28
IFN-2
CTGTGGGTCT GAGGCAGAT

SEQ ID NO. 29
IFN-3b
AACTCCTCTT GAGGAAAGCC AAAATC

SEQ ID NO. 30
IFN-4
ATCTGCTGGA TCATCTCGTG C

SEQ ID NO. 31
IFN-8
TGCTCTAGAC TTTTTACTCC TTAGACCTCA AGCTCT

FIG. 8B

SEQ ID NO. 32
Oligo 1. TCACTCGAGG TGAATATCCA AGAAT

SEQ ID NO. 33
Oligo 2. GAGATCGATT TTGGCTGGAC ACTTG

SEQ ID NO. 34
Oligo 3. CACATCGATG TCACAACTTG GGAAT

SEQ ID NO. 35
Oligo 4. TCTAAGCTTC GTCACAGACC GTCCC

FIG. 9

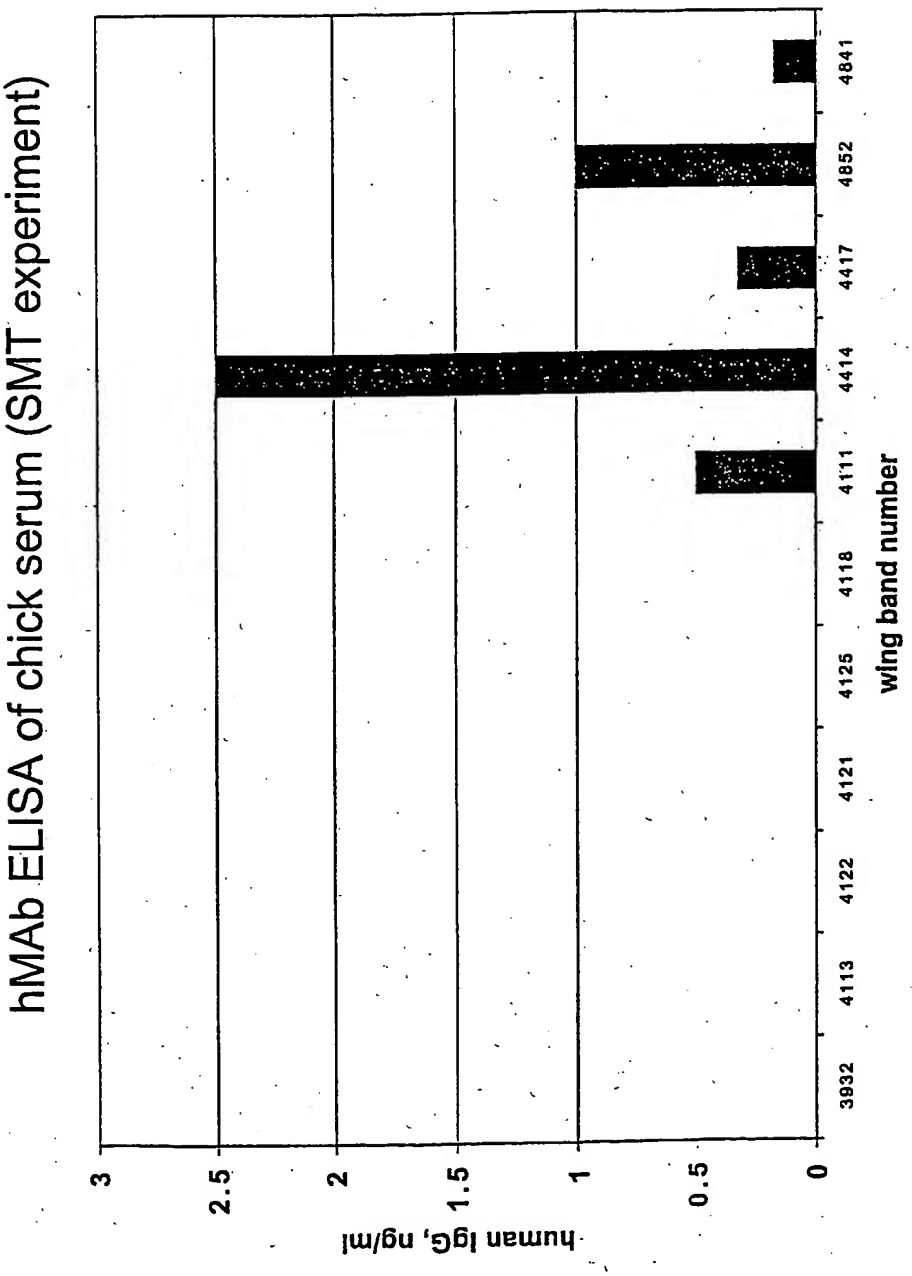
hMAb ELISA of chick serum (SMT experiment)
human IgG, ng/ml
0
0.5
1
1.5
2
2.5
3
3932
4113
4122
4121
4125
4118
4111
4414
4417
4852
4841
wing band number

FIG. 10

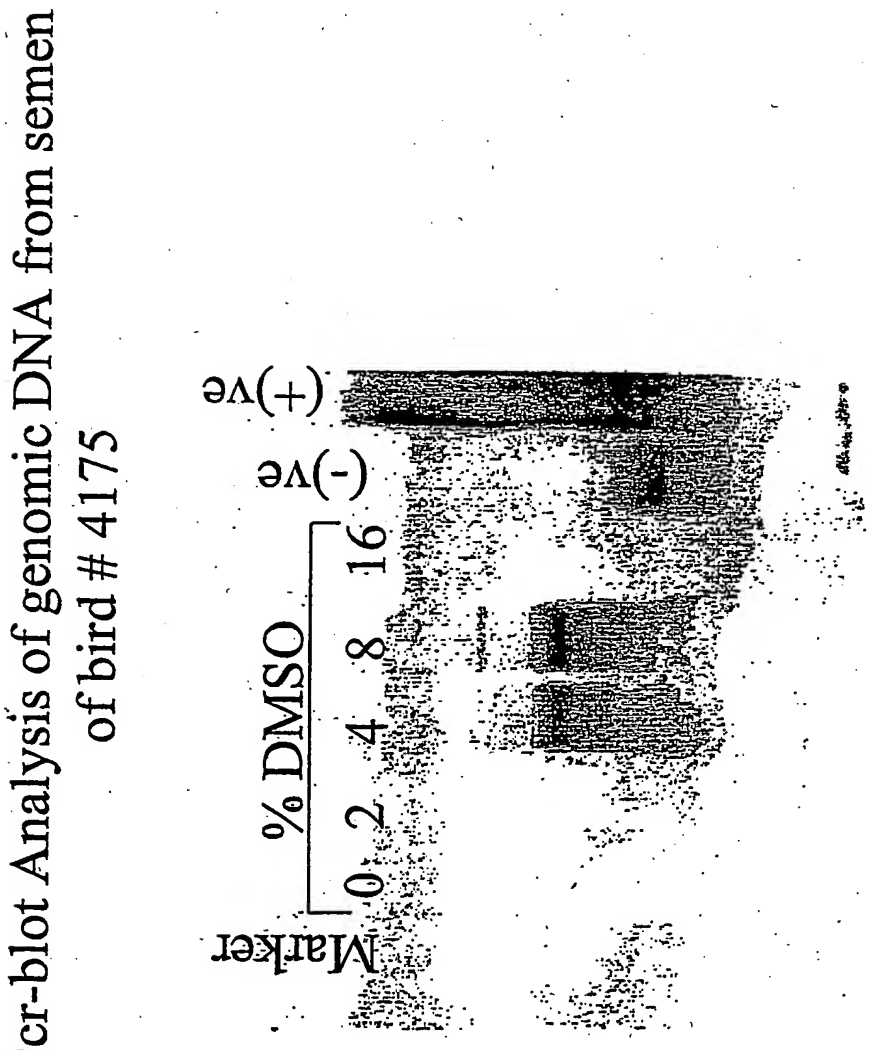
Pcr-blot Analysis of genomic DNA from semen of bird # 4175
Marker
0
2
4
8
16
% DMSO
(-)ve
(+)ve

FIG. 11

SEQUENCE LISTING

<110> AviGenics, Inc.

<120> Production of Transgenic Avians
Using Sperm-mediated Transfection

<130> 11106-021-228

<140> To be assigned
<141> 2002-09-18

<150> 60/324,001
<151> 2001-09-21

<150> 60/323,961
<151> 2001-09-21

<160> 35

<170> PatentIn version 3.1

<210> 1
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> Primer 5pLMAR2

<400> 1
tgccgccttc tttgatattc 20

<210> 2
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> Primer LE-6.1kbrev1

<400> 2
ttggtggtaa ggccttttg 20

<210> 3
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> Primer lys-6.1

<400> 3
ctggcaagct gtcaaaaaca 20

<210> 4
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> Primer LysE1rev

<400> 4
cagctcacat cgtccaaaga 20

<210> 5
<211> 498
<212> DNA
<213> Artificial sequence

<220>
<223> IFNMAGMAX

<220>
<221> misc_feature
<222> (1)..(498)
<223>

<400> 5
tgcgatctgc ctcagaccca cagcctgggc agcaggagga ccctgatgct gctggctcag 60

atgaggagaa tcagcctgtt tagctgcctg aaggataggc acgattttgg ctttcctcaa 120

gaggagtttg gcaaccagtt tcagaaggct gagaccatcc ctgtgctgca cgagatgatc 180

cagcagatct ttaacctgtt tagcaccaag gatagcagcg ctgcttggga tgagaccctg 240

ctggataagt tttacaccga gctgtaccag cagctgaacg atctggaggc ttgcgtgatc 300

cagggcgtgg gcgtgaccga gacccctctg atgaaggagg atagcatcct ggctgtgagg 360

aagtactttc agaggatcac cctgtacctg aaggagaaga agtacagccc ctgcgcttgg 420

gaagtcgtga gggctgagat catgaggagc tttagcctga gcaccaacct gcaagagagc 480

ttgaggtcta aggagtaa 498

<210> 6
<211> 12728
<212> DNA
<213> Gallus gallus

<220>
<221> misc_feature
<222> (1)..(237)
<223> 5prime matrix (scaffold) attachment region (MAR)

<220>
<221> misc_feature

<222> (261)..(1564)
<223> 5prime matrix (scaffold) attachment region (MAR)

<220>
<221> misc_feature
<222> (1565)..(1912)
<223> 5prime matrix (scaffold) attachment region (MAR)

<220>
<221> misc_feature
<222> (1930)..(2012)
<223> 5prime matrix (scaffold) attachment region (MAR)

<220>
<221> misc_feature
<222> (2013)..(2671)
<223> Intrinsically curved DNA

<220>
<221> misc_feature
<222> (5848)..(5934)
<223> Transcription Enhancer

<220>
<221> misc_feature
<222> (9160)..(9325)
<223> Transcription Enhancer

<220>
<221> misc_feature
<222> (9326)..(9626)
<223> Negative Regulatory Element

<220>
<221> misc_feature
<222> (9621)..(9660)
<223> Hormone Response Element

<220>
<221> misc_feature
<222> (9680)..(10060)
<223> Hormone Response Element

<220>
<221> misc_feature
<222> (10576)..(10821)
<223> Chicken CR1 Repeat Sequence

<220>
<221> misc_feature
<222> (10926)..(11193)
<223> Chicken CR1 Repeat Sequence

<220>
<221> misc_feature
<222> (11424)..(11938)
<223> Lysozyme Proximal Promoter and Lysozyme Signal Peptide

<220>
<221> misc_feature
<222> (11946)..(12443)
<223> Human Interferon alpha 2d encoding region codon optimized for exp
ression in chicken cells (IFNMAGMAX)

<220>
<221> polyA_signal
<222> (12444)..(12728)
<223>

<400> 6
tgccgccttc tttgatattc actctgttgt atttcatctc ttcttgccga tgaaaggata 60

taacagtctg tataacagtc tgtgaggaaa tacttggtat ttcttctgat cagtgttttt 120

ataagtaatg ttgaatattg gataaggctg tgtgtccttt gtcttgggag acaaagccca 180

cagcaggtgg tggttggggt ggtggcagct cagtgacagg agaggttttt ttgcctgttt 240

ttttttttt ttttttttt aagtaaggtg ttctttttc ttagtaaatt ttctactgga 300

ctgtatgttt tgacaggtca gaaacatttc ttcaaaagaa gaacctttg gaaactgtac 360

agccctttc tttcattccc tttttgcttt ctgtgccaat gcctttggtt ctgattgcat 420

tatggaaaac gttgatcgga acttgaggtt tttatttata gtgtggcttg aaagcttgga 480

tagctgttgt tacacgagat accttattaa gtttaggcca gcttgatgct ttatttttc 540

cctttgaagt agtgagcgtt ctctggtttt tttcctttga aactggtgag gcttagattt 600

ttctaatggg attttttacc tgatgatcta gttgcatacc caaatgcttg taaatgtttt 660

cctagttaac atgttgataa cttcggattt acatgttgta tatacttgtc atctgtgttt 720

ctagtaaaaa tatatggcat ttatagaaat acgtaattcc tgatttcctt ttttttatc 780

tctatgctct gtgtgtacag gtcaaacaga cttcactcct atttttattt atagaatttt 840

atatgcagtc tgtcgttggt tcttgtgttg taaggataca gccttaaatt tcctagagcg 900

atgctcagta aggcgggttg tcacatgggt tcaaatgtaa aacgggcacg tttggctgct 960

gccttcccga gatccaggac actaaactgc ttctgcactg aggtataaat cgcttcagat 1020

cccagggaag tgcagatcca cgtgcatatt cttaaagaag aatgaatact ttctaaaata 1080

ttttggcata ggaagcaagc tgcatggatt tgtttgggac ttaaattatt ttggtaacgg 1140

agtgcatagg ttttaaacac agttgcagca tgctaacgag tcacagcgtt tatgcagaag 1200

tgatgcctgg atgcctgttg cagctgttta cggcactgcc ttgcagtgag cattgcagat 1260

aggggtgggg tgctttgtgt cgtgttccca cacgctgcca cacagccacc tcccggaaca 1320

catctcacct gctgggtact tttcaaacca tcttagcagt agtagatgag ttactatgaa 1380

acagagaagt tcctcagttg gatattctca tgggatgtct tttttcccat gttgggcaaa 1440

gtatgataaa gcatctctat ttgtaaatta tgcacttgtt agttcctgaa tcctttctat 1500

agcaccactt attgcagcag gtgtaggctc tggtgtggcc tgtgtctgtg cttcaatctt 1560

ttaaagcttc tttggaaata cactgacttg attgaagtct cttgaagata gtaaacagta 1620

cttacctttg atcccaatga aatcgagcat ttcagttgta aaagaattcc gcctattcat 1680

accatgtaat gtaattttac acccccagtg ctgacacttt ggaatatatt caagtaatag 1740

actttggcct caccctcttg tgtactgtat tttgtaatag aaaatatttt aaactgtgca 1800

tatgattatt acattatgaa agagacattc tgctgatctt caaatgtaag aaaatgagga 1860

gtgcgtgtgc ttttataaat acaagtgatt gcaaattagt gcaggtgtcc ttaaaaaaaa 1920

aaaaaaaaag taatataaaa aggaccaggt gttttacaag tgaaatacat tcctatttgg 1980

taaacagtta catttttatg aagattacca gcgctgctga ctttctaaac ataaggctgt 2040

attgtcttcc tgtaccattg catttcctca ttcccaattt gcacaaggat gtctgggtaa 2100

actattcaag aaatggcttt gaaatacagc atgggagctt gtctgagttg gaatgcagag 2160

ttgcactgca aaatgtcagg aaatggatgt ctctcagaat gcccaactcc aaaggatttt 2220

atatgtgtat atagtaagca gtttcctgat tccagcaggc caaagagtct gctgaatgtt 2280

gtgttgccgg agacctgtat ttctcaacaa ggtaagatgg tatcctagca actgcggatt 2340

ttaatacatt ttcagcagaa gtacttagtt aatctctacc tttagggatc gtttcatcat 2400

ttttagatgt tatacttgaa atactgcata acttttagct ttcatgggtt ccttttttttc 2460

agcctttagg agactgttaa gcaatttgct gtccaacttt tgtgttggtc ttaaactgca 2520

atagtagttt accttgtatt gaagaaataa agaccatttt tatattaaaa aatacttttg 2580

tctgtcttca ttttgacttg tctgatatcc ttgcagtgcc cattatgtca gttctgtcag 2640

atattcagac atcaaaactt aacgtgagct cagtggagtt acagctgcgg ttttgatgct 2700

gttattattt ctgaaactag aaatgatgtt gtcttcatct gctcatcaaa cacttcatgc 2760

agagtgtaag gctagtgaga aatgcataca tttattgata ctttttttaaa gtcaactttt 2820
tatcagattt ttttttcatt tggaaatata ttgttttcta gactgcatag cttctgaatc 2880
tgaaatgcag tctgattggc atgaagaagc acagcactct tcatcttact taaacttcat 2940
tttggaatga aggaagttaa gcaagggcac aggtccatga aatagagaca gtgcgctcag 3000
gagaaagtga acctggattt ctttggctag tgttctaaat ctgtagtgag gaaagtaaca 3060
cccgattcct tgaaagggct ccagctttaa tgcttccaaa ttgaaggtgg caggcaactt 3120
ggccactggt tatttactgc attatgtctc agtttcgcag ctaacctggc ttctccacta 3180
ttgagcatgg actatagcct ggcttcagag gccaggtgaa ggttgggatg ggtggaagga 3240
gtgctgggct gtggctgggg ggactgtggg gactccaagc tgagcttggg gtgggcagca 3300
cagggaaaag tgtgggtaac tatttttaag tactgtgttg caaacgtctc atctgcaaat 3360
acgtagggtg tgtactctcg aagattaaca gtgtgggttc agtaatatat ggatgaattc 3420
acagtggaag cattcaaggg tagatcatct aacgacacca gatcatcaag ctatgattgg 3480
aagcggtatc agaagagcga ggaaggtaag cagtcttcat atgttttccc tccacgtaaa 3540
gcagtctggg aaagtagcac cccttgagca gagacaagga aataattcag gagcatgtgc 3600
taggagaact ttcttgctga attctacttg caagagcttt gatgcctggc ttctggtgcc 3660
ttctgcagca cctgcaaggc ccagagcctg tggtgagctg gagggaaaga ttctgctcaa 3720
gtccaagctt cagcaggtca ttgtctttgc ttcttccccc agcactgtgc agcagagtgg 3780
aactgatgtc gaagcctcct gtccactacc tgttgctgca ggcagactgc tctcagaaaa 3840
agagagctaa ctctatgcca tagtctgaag gtaaaatggg ttttaaaaaa gaaaacacaa 3900
aggcaaaacc ggctgcccca tgagaagaaa gcagtggtaa acatggtaga aaaggtgcag 3960
aagccccccag gcagtgtgac aggcccctcc tgccacctag aggcgggaac aagcttccct 4020
gcctagggct ctgcccgcga agtgcgtgtt tctttggtgg gttttgtttg gcgtttggtt 4080
ttgagattta gacacaaggg aagcctgaaa ggaggtgttg ggcactattt tggtttgtaa 4140
agcctgtact tcaaatatat attttgtgag ggagtgtagc gaattggcca atttaaaata 4200
aagttgcaag agattgaagg ctgagtagtt gagagggtaa cacgtttaat gagatcttct 4260
gaaactactg cttctaaaca cttgtttgag tggtgagacc ttggataggt gagtgctctt 4320
gttacatgtc tgatgcactt gcttgtcctt ttccatccac atccatgcat tccacatcca 4380
cgcatttgtc acttatccca tatctgtcat atctgacata cctgtctctt cgtcacttgg 4440
tcagaagaaa cagatgtgat aatccccagc cgccccaagt ttgagaagat ggcagttgct 4500

tctttccctt tttcctgcta agtaaggatt ttctcctggc tttgacacct cacgaaatag 4560
tcttcctgcc ttacattctg ggcattattt caaatatctt tggagtgcgc tgctctcaag 4620
tttgtgtctt cctactctta gagtgaatgc tcttagagtg aaagagaagg aagagaagat 4680
gttggccgca gttctctgat gaacacacct ctgaataatg gccaaaggtg ggtgggtttc 4740
tctgaggaac gggcagcgtt tgcctctgaa agcaaggagc tctgcggagt tgcagttatt 4800
ttgcaactga tggtggaact ggtgcttaaa gcagattccc taggttccct gctacttctt 4860
ttccttcttg gcagtcagtt tatttctgac agacaaacag ccacccccac tgcaggctta 4920
gaaagtatgt ggctctgcct gggtgtgtta cagctctgcc ctggtgaaag gggattaaaa 4980
cgggcaccat tcatcccaaa caggatcctc attcatggat caagctgtaa ggaacttggg 5040
ctccaacctc aaaacattaa ttggagtacg aatgtaatta aaactgcatt ctcgcattcc 5100
taagtcattt agtctggact ctgcagcatg taggtcggca gctcccactt tctcaaagac 5160
cactgatgga ggagtagtaa aaatggagac cgattcagaa caaccaacgg agtgttgccg 5220
aagaaactga tggaaataat gcatgaattg tgtggtggac atttttttta aatacataaa 5280
ctacttcaaa tgaggtcgga gaaggtcagt gttttattag cagccataaa accaggtgag 5340
cgagtaccat ttttctctac aagaaaaacg attctgagct ctgcgtaagt ataagttctc 5400
catagcggct gaagctcccc cctggctgcc tgccatctca gctggagtgc agtgccattt 5460
ccttggggtt tctctcacag cagtaatggg acaatacttc acaaaaattc tttcttttcc 5520
tgtcatgtgg gatccctact gtgccctcct ggttttacgt taccccctga ctgttccatt 5580
cagcggtttg gaaagagaaa aagaatttgg aaataaaaca tgtctacgtt atcacctcct 5640
ccagcatttt ggtttttaat tatgtcaata actggcttag atttggaaat gagagggggt 5700
tgggtgtatt accgaggaac aaaggaaggc ttatataaac tcaagtcttt tatttagaga 5760
actggcaagc tgtcaaaaac aaaaaggcct taccaccaaa ttaagtgaat agccgctata 5820
gccagcaggg ccagcacgag ggatggtgca ctgctggcac tatgccacgg cctgcttgtg 5880
actctgagag caactgcttt ggaaatgaca gcacttggtg caatttcctt tgtttcagaa 5940
tgcgtagagc gtgtgcttgg cgacagtttt tctagttagg ccacttcttt tttccttctc 6000
tcctcattct cctaagcatg tctccatgct ggtaatccca gtcaagtgaa cgttcaaaca 6060
atgaatccat cactgtagga ttctcgtggt gatcaaatct ttgtgtgagg tctataaaat 6120
atggaagctt atttattttt cgttcttcca tatcagtctt ctctatgaca attcacatcc 6180
accacagcaa attaaaggtg aaggaggctg gtgggatgaa gagggtcttc tagctttacg 6240

ttcttccttg caaggccaca ggaaaatgct gagagctgta gaatacagcc tggggtaaga 6300
agttcagtct cctgctggga cagctaaccg catcttataa ccccttctga gactcatctt 6360
aggaccaaat agggtctatc tggggttttt gttcctgctg ttcctcctgg aaggctatct 6420
cactatttca ctgctcccac ggttacaaac caaagataca gcctgaattt tttctaggcc 6480
acattacata aatttgacct ggtaccaata ttgttctcta tatagttatt tccttcccca 6540
ctgtgtttaa ccccttaagg cattcagaac aactagaatc atagaatggt ttggattgga 6600
aggggcctta aacatcatcc atttccaacc ctctgccatg ggctgcttgc cacccactgg 6660
ctcaggctgc ccagggcccc atccagcctg gccttgagca cctccaggga tggggcaccc 6720
acagcttctc tgggcagcct gtgccaacac ctcaccactc tctgggtaaa gaattctctt 6780
ttaacatcta atctaaatct cttctctttt agtttaaagc cattcctctt tttcccgttg 6840
ctatctgtcc aagaaatgtg tattggtctc cctcctgctt ataagcagga agtactggaa 6900
ggctgcagtg aggtctcccc acagccttct cttctccagg ctgaacaagc ccagctcctt 6960
cagcctgtct tcgtaggaga tcatcttagt ggccctcctc tggacccatt ccaacagttc 7020
cacggctttc ttgtggagcc ccaggtctgg atgcagtact tcagatgggg ccttacaaag 7080
gcagagcaga tggggacaat cgcttacccc tccctgctgg ctgcccctgt tttgatgcag 7140
cccagggtac tgttggcctt tcaggctccc agacccctg ctgatttgtg tcaagctttt 7200
catccaccag aacccacgct tcctggttaa tacttctgcc ctcacttctg taagcttgtt 7260
tcaggagact tccattcttt aggacagact gtgttacacc tacctgccct attcttgcat 7320
atatacattt cagttcatgt ttcctgtaac aggacagaat atgtattcct ctaacaaaaa 7380
tacatgcaga attcctagtg ccatctcagt agggttttca tggcagtatt agcacatagt 7440
caatttgctg caagtacctt ccaagctgcg gcctcccata aatcctgtat ttgggatcag 7500
ttacctttg gggtaagctt ttgtatctgc agagaccctg ggggttctga tgtgcttcag 7560
ctctgctctg ttctgactgc accattttct agatcaccca gttgttcctg tacaacttcc 7620
ttgtcctcca tcctttccca gcttgtatct ttgacaaata caggcctatt tttgtgtttg 7680
cttcagcagc catttaattc ttcagtgtca tcttgttctg ttgatgccac tggaacagga 7740
ttttcagcag tcttgcaaag aacatctagc tgaaaacttt ctgccattca atattcttac 7800
cagttcttct tgtttgaggt gagccataaa ttactagaac ttcgtcactg acaagtttat 7860
gcattttatt acttctatta tgtacttact ttgacataac acagacacgc acatattttg 7920
ctgggatttc cacagtgtct ctgtgtcctt cacatggttt tactgtcata cttccgttat 7980

aaccttggca atctgcccag ctgcccatca caagaaaaga gattcctttt ttattacttc 8040
tcttcagcca ataaacaaaa tgtgagaagc ccaaacaaga acttgtgggg caggctgcca 8100
tcaagggaga gacagctgaa gggttgtgta gctcaataga attaagaaat aataaagctg 8160
tgtcagacag ttttgcctga tttatacagg cacgccccaa gccagagagg ctgtctgcca 8220
aggccacctt gcagtccttg gtttgtaaga taagtcatag gtaacttttc tggtgaattg 8280
cgtggagaat catgatggca gttcttgctg tttactatgg taagatgcta aaataggaga 8340
cagcaaagta acacttgctg ctgtaggtgc tctgctatcc agacagcgat ggcactcgca 8400
caccaagatg agggatgctc ccagctgacg gatgctgggg cagtaacagt gggtcccatg 8460
ctgcctgctc attagcatca cctcagccct caccagccca tcagaaggat catcccaagc 8520
tgaggaaagt tgctcatctt cttcacatca tcaaaccttt ggcctgactg atgcctcccg 8580
gatgcttaaa tgtggtcact gacatcttta tttttctatg atttcaagtc agaacctccg 8640
gatcaggagg gaacacatag tgggaatgta ccctcagctc caaggccaga tcttccttca 8700
atgatcatgc atgctactta ggaaggtgtg tgtgtgtgaa tgtagaattg cctttgttat 8760
tttttcttcc tgctgtcagg aacattttga ataccagaga aaaagaaaag tgctcttctt 8820
ggcatgggag gagttgtcac acttgcaaaa taaaggatgc agtcccaaat gttcataatc 8880
tcagggtctg aaggaggatc agaaactgtg tatacaattt caggcttctc tgaatgcagc 8940
ttttgaaagc tgttcctggc cgaggcagta ctagtcagaa ccctcggaaa caggaacaaa 9000
tgtcttcaag gtgcagcagg aggaaacacc ttgcccatca tgaaagtgaa taaccactgc 9060
cgctgaagga atccagctcc tgtttgagca ggtgctgcac actcccacac tgaaacaaca 9120
gttcattttt ataggacttc caggaaggat cttcttctta agcttcttaa ttatggtaca 9180
tctccagttg gcagatgact atgactactg acaggagaat gaggaactag ctgggaatat 9240
ttctgtttga ccaccatgga gtcacccatt tctttactgg tatttggaaa taataattct 9300
gaattgcaaa gcaggagtta gcgaagatct tcatttcttc catgttggtg acagcacagt 9360
tctggctatg aaagtctgct tacaaggaag aggataaaaa tcatagggat aataaatcta 9420
agtttgaaga caatgaggtt ttagctgcat ttgacatgaa gaaattgaga cctctactgg 9480
atagctatgg tatttacgtg tctttttgct tagttactta ttgaccccag ctgaggtcaa 9540
gtatgaactc aggtctctcg ggctactggc atggattgat tacatacaac tgtaatttta 9600
gcagtgattt agggtttatg agtacttttg cagtaaatca tagggttagt aatgttaatc 9660
tcagggaaaa aaaaaaaaag ccaaccctga cagacatccc agctcaggtg gaaatcaagg 9720

atcacagctc agtgcggtcc cagagaacac agggactctt ctcttaggac ctttatgtac 9780
agggcctcaa gataactgat gttagtcaga agactttcca ttctggccac agttcagctg 9840
aggcaatcct ggaattttct ctccgctgca cagttccagt catcccagtt tgtacagttc 9900
tggcactttt tgggtcaggc cgtgatccaa ggagcagaag ttccagctat ggtcagggag 9960
tgcctgaccg tcccaactca ctgcactcaa acaaaggcga aaccacaaga gtggcttttg 10020
ttgaaattgc agtgtggccc agaggggctg caccagtact ggattgacca cgaggcaaca 10080
ttaatcctca gcaagtgcaa tttgcagcca ttaaattgaa ctaactgata ctacaatgca 10140
atcagtatca acaagtggtt tggcttggaa gatggagtct aggggctcta caggagtagc 10200
tactctctaa tggagttgca ttttgaagca ggacactgtg aaaagctggc ctcctaaaga 10260
ggctgctaaa cattagggtc aattttccag tgcactttct gaagtgtctg cagttcccca 10320
tgcaaagctg cccaaacata gcacttccaa ttgaatacaa ttatatgcag gcgtactgct 10380
tcttgccagc actgtccttc tcaaatgaac tcaacaaaca atttcaaagt ctagtagaaa 10440
gtaacaagct ttgaatgtca ttaaaaagta tatctgcttt cagtagttca gcttatttat 10500
gcccactaga aacatcttgt acaagctgaa cactggggct ccagattagt ggtaaaacct 10560
actttataca atcatagaat catagaatgg cctgggttgg aagggacccc aaggatcatg 10620
aagatccaac acccccgcca caggcagggc caccaacctc cagatctggt actagaccag 10680
gcagcccagg gctccatcca acctggccat gaacacctcc agggatggag catccacaac 10740
ctctctgggc agcctgtgcc agcacctcac caccctctct gtgaagaact tttccctgac 10800
atccaatcta agccttccct ccttgaggtt agatccactc ccccttgtgc tatcactgtc 10860
tactcttgta aaaagttgat tctcctcctt tttggaaggt tgcaatgagg tctccttgca 10920
gccttcttct cttctgcagg atgaacaagc ccagctccct cagcctgtct ttataggaga 10980
ggtgctccag ccctctgatc atctttgtgg ccctcctctg gacccgctcc aagagctcca 11040
catctttcct gtactggggg ccccaggcct gaatgcagta ctccagatgg ggcctcaaaa 11100
gagcagagta aagagggaca atcaccttcc tcaccctgct ggccagccct cttctgatgg 11160
agccctggat acaactggct ttctgagctg caacttctcc ttatcagttc cactattaaa 11220
acaggaacaa tacaacaggt gctgatggcc agtgcagagt ttttcacact tcttcatttc 11280
ggtagatctt agatgaggaa cgttgaagtt gtgcttctgc gtgtgcttct tcctcctcaa 11340
atactcctgc ctgatacctc accccacctg ccactgaatg gctccatggc cccctgcagc 11400
cagggccctg atgaacccgg cactgcttca gatgctgttt aatagcacag tatgaccaag 11460

ttgcacctat gaatacacaa acaatgtgtt gcatccttca gcacttgaga agaagagcca 11520

aatttgcatt gtcaggaaat ggtttagtaa ttctgccaat taaaacttgt ttatctacca 11580

tggctgtttt tatggctgtt agtagtggta cactgatgat gaacaatggc tatgcagtaa 11640

aatcaagact gtagatattg caacagacta taaaattcct ctgtggctta gccaatgtgg 11700

tacttcccac attgtataag aaatttggca agtttagagc aatgtttgaa gtgttgggaa 11760

atttctgtat actcaagagg gcgtttttga caactgtaga acagaggaat caaaaggggg 11820

tgggaggaag ttaaaagaag aggcaggtgc aagagagctt gcagtcccgc tgtgtgtacg 11880

acactggcaa catgaggtct ttgctaatct tggtgctttg cttcctgccc ctggctgcct 11940

tagggtgcga tctgcctcag acccacagcc tgggcagcag gaggaccctg atgctgctgg 12000

ctcagatgag gagaatcagc ctgtttagct gcctgaagga taggcacgat tttggctttc 12060

ctcaagagga gtttggcaac cagtttcaga aggctgagac catccctgtg ctgcacgaga 12120

tgatccagca gatctttaac ctgtttagca ccaaggatag cagcgctgct tgggatgaga 12180

ccctgctgga taagttttac accgagctgt accagcagct gaacgatctg gaggcttgcg 12240

tgatccaggg cgtgggcgtg accgagaccc ctctgatgaa ggaggatagc atcctggctg 12300

tgaggaagta ctttcagagg atcaccctgt acctgaagga gaagaagtac agcccctgcg 12360

cttgggaagt cgtgagggct gagatcatga ggagctttag cctgagcacc aacctgcaag 12420

agagcttgag gtctaaggag taaaaagtct agagtcgggg cggccggccg cttcgagcag 12480

acatgataag atacattgat gagtttggac aaaccacaac tagaatgcag tgaaaaaaat 12540

gctttatttg tgaaatttgt gatgctattg ctttatttgt aaccattata agctgcaata 12600

aacaagttaa caacaacaat tgcattcatt ttatgtttca ggttcagggg gaggtgtggg 12660

aggtttttta aagcaagtaa aacctctaca aatgtggtaa aatcgataag gatccgtcga 12720

gcggccgc 12728

<210> 7
<211> 11945
<212> DNA
<213> Gallus gallus

<220>
<221> misc_feature
<222> (1)..(237)
<223> 5prime matrix attachment region (MAR)

<220>
<221> misc_feature

<222> (261)..(1564)
<223> 5prime matrix attachment region (MAR)

<220>
<221> misc_feature
<222> (1565)..(1912)
<223> 5prime matrix attachment region (MAR)

<220>
<221> misc_feature
<222> (1930)..(2012)
<223> 5prime matrix attachment region (MAR)

<220>
<221> misc_feature
<222> (2013)..(2671)
<223> Intrinsically Curved DNA

<220>
<221> misc_feature
<222> (5848)..(5934)
<223> Transcription Enhancer

<220>
<221> misc_feature
<222> (9160)..(9325)
<223> Transcription Enhancer

<220>
<221> misc_feature
<222> (9326)..(9626)
<223> Negative Regulatory Element

<220>
<221> misc_feature
<222> (9621)..(9660)
<223> Hormone Response Element

<220>
<221> misc_feature
<222> (9680)..(10060)
<223> Hormone Response Element

<220>
<221> misc_feature
<222> (10576)..(10821)
<223> Chicken CR1 Repeat

<220>
<221> misc_feature
<222> (10926)..(11193)
<223> Chicken CR1 Repeat

<220>
<221> misc_feature
<222> (11424)..(11938)
<223> Proximal promoter and lysozyme signal peptide

<400> 7
tgccgccttc tttgatattc actctgttgt atttcatctc ttcttgccga tgaaaggata 60

taacagtctg tataacagtc tgtgaggaaa tacttggtat ttcttctgat cagtgttttt 120

ataagtaatg ttgaatattg gataaggctg tgtgtccttt gtcttgggag acaaagccca 180

cagcaggtgg tggttggggt ggtggcagct cagtgacagg agaggttttt ttgcctgttt 240

ttttttttt ttttttttt aagtaaggtg ttctttttc ttagtaaatt ttctactgga 300

ctgtatgttt tgacaggtca gaaacatttc ttcaaaagaa gaacctttg gaaactgtac 360

agccctttc tttcattccc tttttgcttt ctgtgccaat gcctttggtt ctgattgcat 420

tatggaaaac gttgatcgga acttgaggtt tttatttata gtgtggcttg aaagcttgga 480

tagctgttgt tacacgagat accttattaa gtttaggcca gcttgatgct ttatttttc 540

cctttgaagt agtgagcgtt ctctggtttt tttcctttga aactggtgag gcttagattt 600

ttctaatggg attttttacc tgatgatcta gttgcatacc caaatgcttg taaatgtttt 660

cctagttaac atgttgataa cttcggattt acatgttgta tatacttgtc atctgtgttt 720

ctagtaaaaa tatatggcat ttatagaaat acgtaattcc tgatttcctt tttttttatc 780

tctatgctct gtgtgtacag gtcaaacaga cttcactcct atttttattt atagaatttt 840

atatgcagtc tgtcgttggt tcttgtgttg taaggataca gccttaaatt tcctagagcg 900

atgctcagta aggcgggttg tcacatgggt tcaaatgtaa aacgggcacg tttggctgct 960

gccttcccga gatccaggac actaaactgc ttctgcactg aggtataaat cgcttcagat 1020

cccagggaag tgcagatcca cgtgcatatt cttaaagaag aatgaatact ttctaaaata 1080

ttttggcata ggaagcaagc tgcatggatt tgtttgggac ttaaattatt ttggtaacgg 1140

agtgcatagg ttttaaacac agttgcagca tgctaacgag tcacagcgtt tatgcagaag 1200

tgatgcctgg atgcctgttg cagctgttta cggcactgcc ttgcagtgag cattgcagat 1260

aggggtgggg tgctttgtgt cgtgttccca cacgctgcca cacagccacc tcccggaaca 1320

catctcacct gctgggtact tttcaaacca tcttagcagt agtagatgag ttactatgaa 1380

acagagaagt tcctcagttg gatattctca tgggatgtct tttttcccat gttgggcaaa 1440
gtatgataaa gcatctctat ttgtaaatta tgcacttgtt agttcctgaa tcctttctat 1500
agcaccactt attgcagcag gtgtaggctc tggtgtggcc tgtgtctgtg cttcaatctt 1560
ttaaagcttc tttggaaata cactgacttg attgaagtct cttgaagata gtaaacagta 1620
cttacctttg atcccaatga aatcgagcat ttcagttgta aaagaattcc gcctattcat 1680
accatgtaat gtaattttac acccccagtg ctgacacttt ggaatatatt caagtaatag 1740
actttggcct caccctcttg tgtactgtat tttgtaatag aaaatatttt aaactgtgca 1800
tatgattatt acattatgaa agagacattc tgctgatctt caaatgtaag aaaatgagga 1860
gtgcgtgtgc ttttataaat acaagtgatt gcaaattagt gcaggtgtcc ttaaaaaaaa 1920
aaaaaaaaag taatataaaa aggaccaggt gttttacaag tgaaatacat tcctatttgg 1980
taaacagtta catttttatg aagattacca gcgctgctga ctttctaaac ataaggctgt 2040
attgtcttcc tgtaccattg catttcctca ttcccaattt gcacaaggat gtctgggtaa 2100
actattcaag aaatggcttt gaaatacagc atgggagctt gtctgagttg gaatgcagag 2160
ttgcactgca aaatgtcagg aaatggatgt ctctcagaat gcccaactcc aaaggatttt 2220
atatgtgtat atagtaagca gtttcctgat tccagcaggc caaagagtct gctgaatgtt 2280
gtgttgccgg agacctgtat ttctcaacaa ggtaagatgg tatcctagca actgcggatt 2340
ttaatacatt ttcagcagaa gtacttagtt aatctctacc tttagggatc gtttcatcat 2400
ttttagatgt tatacttgaa atactgcata acttttagct ttcatgggtt cctttttttc 2460
agcctttagg agactgttaa gcaatttgct gtccaacttt tgtgttggtc ttaaactgca 2520
atagtagttt accttgtatt gaagaaataa agaccatttt tatattaaaa aatacttttg 2580
tctgtcttca ttttgacttg tctgatatcc ttgcagtgcc cattatgtca gttctgtcag 2640
atattcagac atcaaaactt aacgtgagct cagtggagtt acagctgcgg ttttgatgct 2700
gttattattt ctgaaactag aaatgatgtt gtcttcatct gctcatcaaa cacttcatgc 2760
agagtgtaag gctagtgaga aatgcataca tttattgata ctttttaaa gtcaactttt 2820
tatcagattt ttttttcatt tggaaatata ttgttttcta gactgcatag cttctgaatc 2880
tgaaatgcag tctgattggc atgaagaagc acagcactct tcatcttact taaacttcat 2940
tttggaatga aggaagttaa gcaagggcac aggtccatga aatagagaca gtgcgctcag 3000
gagaaagtga acctggattt ctttggctag tgttctaaat ctgtagtgag gaaagtaaca 3060
cccgattcct tgaaagggct ccagctttaa tgcttccaaa ttgaaggtgg caggcaactt 3120

ggccactggt tatttactgc attatgtctc agtttcgcag ctaacctggc ttctccacta 3180
ttgagcatgg actatagcct ggcttcagag gccaggtgaa ggttgggatg ggtggaagga 3240
gtgctgggct gtggctgggg ggactgtggg gactccaagc tgagcttggg gtgggcagca 3300
cagggaaaag tgtgggtaac tatttttaag tactgtgttg caaacgtctc atctgcaaat 3360
acgtagggtg tgtactctcg aagattaaca gtgtgggttc agtaatatat ggatgaattc 3420
acagtggaag cattcaaggg tagatcatct aacgacacca gatcatcaag ctatgattgg 3480
aagcggtatc agaagagcga ggaaggtaag cagtcttcat atgttttccc tccacgtaaa 3540
gcagtctggg aaagtagcac cccttgagca gagacaagga aataattcag gagcatgtgc 3600
taggagaact ttcttgctga attctacttg caagagcttt gatgcctggc ttctggtgcc 3660
ttctgcagca cctgcaaggc ccagagcctg tggtgagctg gagggaaaga ttctgctcaa 3720
gtccaagctt cagcaggtca ttgtctttgc ttcttccccc agcactgtgc agcagagtgg 3780
aactgatgtc gaagcctcct gtccactacc tgttgctgca ggcagactgc tctcagaaaa 3840
agagagctaa ctctatgcca tagtctgaag gtaaaatggg ttttaaaaaa gaaaacacaa 3900
aggcaaaacc ggctgcccca tgagaagaaa gcagtggtaa acatggtaga aaaggtgcag 3960
aagcccccag gcagtgtgac aggcccctcc tgccacctag aggcgggaac aagcttccct 4020
gcctagggct ctgcccgcga agtgcgtgtt tctttggtgg gttttgtttg gcgtttggtt 4080
ttgagattta gacacaaggg aagcctgaaa ggaggtgttg ggcactattt tggtttgtaa 4140
agcctgtact tcaaatatat attttgtgag ggagtgtagc gaattggcca atttaaaata 4200
aagttgcaag agattgaagg ctgagtagtt gagagggtaa cacgtttaat gagatcttct 4260
gaaactactg cttctaaaca cttgtttgag tggtgagacc ttggataggt gagtgctctt 4320
gttacatgtc tgatgcactt gcttgtcctt ttccatccac atccatgcat tccacatcca 4380
cgcatttgtc acttatccca tatctgtcat atctgacata cctgtctctt cgtcacttgg 4440
tcagaagaaa cagatgtgat aatccccagc cgccccaagt ttgagaagat ggcagttgct 4500
tctttccctt tttcctgcta agtaaggatt ttctcctggc tttgacacct cacgaaatag 4560
tcttcctgcc ttacattctg ggcattattt caaatatctt tggagtgcgc tgctctcaag 4620
tttgtgtctt cctactctta gagtgaatgc tcttagagtg aaagagaagg aagagaagat 4680
gttggccgca gttctctgat gaacacacct ctgaataatg gccaaaggtg ggtgggtttc 4740
tctgaggaac gggcagcgtt tgcctctgaa agcaaggagc tctgcggagt tgcagttatt 4800
ttgcaactga tggtggaact ggtgcttaaa gcagattccc taggttccct gctacttctt 4860

ttccttcttg gcagtcagtt tatttctgac agacaaacag ccacccccac tgcaggctta 4920
gaaagtatgt ggctctgcct gggtgtgtta cagctctgcc ctggtgaaag gggattaaaa 4980
cgggcaccat tcatcccaaa caggatcctc attcatggat caagctgtaa ggaacttggg 5040
ctccaacctc aaaacattaa ttggagtacg aatgtaatta aaactgcatt ctcgcattcc 5100
taagtcattt agtctggact ctgcagcatg taggtcggca gctcccactt tctcaaagac 5160
cactgatgga ggagtagtaa aaatggagac cgattcagaa caaccaacgg agtgttgccg 5220
aagaaactga tggaaataat gcatgaattg tgtggtggac attttttttа aatacataaa 5280
ctacttcaaa tgaggtcgga gaaggtcagt gttttattag cagccataaa accaggtgag 5340
cgagtaccat ttttctctac aagaaaaacg attctgagct ctgcgtaagt ataagttctc 5400
catagcggct gaagctcccc cctggctgcc tgccatctca gctggagtgc agtgccattt 5460
ccttggggtt tctctcacag cagtaatggg acaatacttc acaaaaattc tttcttttcc 5520
tgtcatgtgg gatccctact gtgccctcct ggttttacgt taccccctga ctgttccatt 5580
cagcggtttg gaaagagaaa aagaatttgg aaataaaaca tgtctacgtt atcacctcct 5640
ccagcatttt ggtttttaat tatgtcaata actggcttag atttggaaat gagagggggt 5700
tgggtgtatt accgaggaac aaaggaaggc ttatataaac tcaagtcttt tatttagaga 5760
actggcaagc tgtcaaaaac aaaaaggcct taccaccaaa ttaagtgaat agccgctata 5820
gccagcaggg ccagcacgag ggatggtgca ctgctggcac tatgccacgg cctgcttgtg 5880
actctgagag caactgcttt ggaaatgaca gcacttggtg caatttcctt tgtttcagaa 5940
tgcgtagagc gtgtgcttgg cgacagtttt tctagttagg ccacttcttt tttccttctc 6000
tcctcattct cctaagcatg tctccatgct ggtaatccca gtcaagtgaa cgttcaaaca 6060
atgaatccat cactgtagga ttctcgtggt gatcaaatct ttgtgtgagg tctataaaat 6120
atggaagctt atttatttтt cgttcttcca tatcagtctt ctctatgaca attcacatcc 6180
accacagcaa attaaaggtg aaggaggctg gtgggatgaa gagggtcttc tagctttacg 6240
ttcttccttg caaggccaca ggaaaatgct gagagctgta gaatacagcc tggggtaaga 6300
agttcagtct cctgctggga cagctaaccg catcttataa ccccttctga gactcatctt 6360
aggaccaaat agggtctatc tggggttttt gttcctgctg ttcctcctgg aaggctatct 6420
cactatttca ctgctcccac ggttacaaac caaagataca gcctgaattt tttctaggcc 6480
acattacata aatttgacct ggtaccaata ttgttctcta tatagttatt tccttcccca 6540
ctgtgtttaa ccccttaagg cattcagaac aactagaatc atagaatggt ttggattgga 6600

aggggcctta aacatcatcc atttccaacc ctctgccatg ggctgcttgc cacccactgg 6660
ctcaggctgc ccagggcccc atccagcctg gccttgagca cctccaggga tggggcaccc 6720
acagcttctc tgggcagcct gtgccaacac ctcaccactc tctgggtaaa gaattctctt 6780
ttaacatcta atctaaatct cttctctttt agtttaaagc cattcctctt tttcccgttg 6840
ctatctgtcc aagaaatgtg tattggtctc cctcctgctt ataagcagga agtactggaa 6900
ggctgcagtg aggtctcccc acagccttct cttctccagg ctgaacaagc ccagctcctt 6960
cagcctgtct tcgtaggaga tcatcttagt ggccctcctc tggacccatt ccaacagttc 7020
cacggctttc ttgtggagcc ccaggtctgg atgcagtact tcagatgggg ccttacaaag 7080
gcagagcaga tggggacaat cgcttacccc tccctgctgg ctgcccctgt tttgatgcag 7140
cccagggtac tgttggcctt tcaggctccc agacccctg ctgatttgtg tcaagctttt 7200
catccaccag aacccacgct tcctggttaa tacttctgcc ctcacttctg taagcttgtt 7260
tcaggagact tccattcttt aggacagact gtgttacacc tacctgccct attcttgcat 7320
atatacattt cagttcatgt ttcctgtaac aggacagaat atgtattcct ctaacaaaaa 7380
tacatgcaga attcctagtg ccatctcagt agggttttca tggcagtatt agcacatagt 7440
caatttgctg caagtacctt ccaagctgcg gcctcccata aatcctgtat ttgggatcag 7500
ttacctttg gggtaagctt ttgtatctgc agagaccctg ggggttctga tgtgcttcag 7560
ctctgctctg ttctgactgc accattttct agatcaccca gttgttcctg tacaacttcc 7620
ttgtcctcca tcctttccca gcttgtatct ttgacaaata caggcctatt tttgtgtttg 7680
cttcagcagc catttaattc ttcagtgtca tcttgttctg ttgatgccac tggaacagga 7740
ttttcagcag tcttgcaaag aacatctagc tgaaaacttt ctgccattca atattcttac 7800
cagttcttct tgtttgaggt gagccataaa ttactagaac ttcgtcactg acaagtttat 7860
gcattttatt acttctatta tgtacttact ttgacataac acagacacgc acatattttg 7920
ctgggatttc cacagtgtct ctgtgtcctt cacatggttt tactgtcata cttccgttat 7980
aaccttggca atctgcccag ctgcccatca caagaaaaga gattcctttt ttattacttc 8040
tcttcagcca ataaacaaaa tgtgagaagc ccaaacaaga acttgtgggg caggctgcca 8100
tcaagggaga gacagctgaa gggttgtgta gctcaataga attaagaaat aataaagctg 8160
tgtcagacag ttttgcctga tttatacagg cacgccccaa gccagagagg ctgtctgcca 8220
aggccacctt gcagtccttg gtttgtaaga taagtcatag gtaacttttc tggtgaattg 8280
cgtggagaat catgatggca gttcttgctg tttactatgg taagatgcta aaataggaga 8340

cagcaaagta acacttgctg ctgtaggtgc tctgctatcc agacagcgat ggcactcgca 8400
caccaagatg agggatgctc ccagctgacg gatgctgggg cagtaacagt gggtcccatg 8460
ctgcctgctc attagcatca cctcagccct caccagccca tcagaaggat catcccaagc 8520
tgaggaaagt tgctcatctt cttcacatca tcaaaccttt ggcctgactg atgcctcccg 8580
gatgcttaaa tgtggtcact gacatcttta tttttctatg atttcaagtc agaacctccg 8640
gatcaggagg gaacacatag tgggaatgta ccctcagctc caaggccaga tcttccttca 8700
atgatcatgc atgctactta ggaaggtgtg tgtgtgtgaa tgtagaattg cctttgttat 8760
ttttttcttcc tgctgtcagg aacattttga ataccagaga aaaagaaaag tgctcttctt 8820
ggcatgggag gagttgtcac acttgcaaaa taaaggatgc agtcccaaat gttcataatc 8880
tcagggtctg aaggaggatc agaaactgtg tatacaattt caggcttctc tgaatgcagc 8940
ttttgaaagc tgttcctggc cgaggcagta ctagtcagaa ccctcggaaa caggaacaaa 9000
tgtcttcaag gtgcagcagg aggaaacacc ttgcccatca tgaaagtgaa taaccactgc 9060
cgctgaagga atccagctcc tgtttgagca ggtgctgcac actcccacac tgaaacaaca 9120
gttcattttt ataggacttc caggaaggat cttcttctta agcttcttaa ttatggtaca 9180
tctccagttg gcagatgact atgactactg acaggagaat gaggaactag ctgggaatat 9240
ttctgtttga ccaccatgga gtcacccatt tctttactgg tatttggaaa taataattct 9300
gaattgcaaa gcaggagtta gcgaagatct tcatttcttc catgttggtg acagcacagt 9360
tctggctatg aaagtctgct tacaaggaag aggataaaaa tcatagggat aataaatcta 9420
agtttgaaga caatgaggtt ttagctgcat ttgacatgaa gaaattgaga cctctactgg 9480
atagctatgg tatttacgtg tctttttgct tagttactta ttgaccccag ctgaggtcaa 9540
gtatgaactc aggtctctcg ggctactggc atggattgat tacatacaac tgtaatttta 9600
gcagtgattt agggtttatg agtacttttg cagtaaatca tagggttagt aatgttaatc 9660
tcagggaaaa aaaaaaaaag ccaaccctga cagacatccc agctcaggtg gaaatcaagg 9720
atcacagctc agtgcggtcc cagagaacac agggactctt ctcttaggac ctttatgtac 9780
agggcctcaa gataactgat gttagtcaga agactttcca ttctggccac agttcagctg 9840
aggcaatcct ggaattttct ctccgctgca cagttccagt catcccagtt tgtacagttc 9900
tggcactttt tgggtcaggc cgtgatccaa ggagcagaag ttccagctat ggtcagggag 9960
tgcctgaccg tcccaactca ctgcactcaa acaaaggcga aaccacaaga gtggcttttg 10020
ttgaaattgc agtgtggccc agaggggctg caccagtact ggattgacca cgaggcaaca 10080

ttaatcctca gcaagtgcaa tttgcagcca ttaaattgaa ctaactgata ctacaatgca 10140
atcagtatca acaagtggtt tggcttggaa gatggagtct aggggctcta caggagtagc 10200
tactctctaa tggagttgca ttttgaagca ggacactgtg aaaagctggc ctcctaaaga 10260
ggctgctaaa cattagggtc aattttccag tgcactttct gaagtgtctg cagttcccca 10320
tgcaaagctg cccaaacata gcacttccaa ttgaatacaa ttatatgcag gcgtactgct 10380
tcttgccagc actgtccttc tcaaatgaac tcaacaaaca atttcaaagt ctagtagaaa 10440
gtaacaagct ttgaatgtca ttaaaaagta tatctgcttt cagtagttca gcttatttat 10500
gcccactaga aacatcttgt acaagctgaa cactggggct ccagattagt ggtaaaacct 10560
actttataca atcatagaat catagaatgg cctgggttgg aagggacccc aaggatcatg 10620
aagatccaac acccccgcca caggcagggc caccaacctc cagatctggt actagaccag 10680
gcagcccagg gctccatcca acctggccat gaacacctcc agggatggag catccacaac 10740
ctctctgggc agcctgtgcc agcacctcac caccctctct gtgaagaact tttccctgac 10800
atccaatcta agccttccct ccttgaggtt agatccactc ccccttgtgc tatcactgtc 10860
tactcttgta aaaagttgat tctcctcctt tttggaaggt tgcaatgagg tctccttgca 10920
gccttcttct cttctgcagg atgaacaagc ccagctccct cagcctgtct ttataggaga 10980
ggtgctccag ccctctgatc atctttgtgg ccctcctctg gacccgctcc aagagctcca 11040
catctttcct gtactgggggg ccccaggcct gaatgcagta ctccagatgg ggcctcaaaa 11100
gagcagagta aagagggaca atcaccttcc tcaccctgct ggccagccct cttctgatgg 11160
agccctggat acaactggct ttctgagctg caacttctcc ttatcagttc cactattaaa 11220
acaggaacaa tacaacaggt gctgatggcc agtgcagagt ttttcacact tcttcatttc 11280
ggtagatctt agatgaggaa cgttgaagtt gtgcttctgc gtgtgcttct tcctcctcaa 11340
atactcctgc ctgatacctc accccacctg ccactgaatg gctccatggc cccctgcagc 11400
cagggccctg atgaacccgg cactgcttca gatgctgttt aatagcacag tatgaccaag 11460
ttgcacctat gaatacacaa acaatgtgtt gcatccttca gcacttgaga agaagagcca 11520
aatttgcatt gtcaggaaat ggtttagtaa ttctgccaat taaaacttgt ttatctacca 11580
tggctgtttt tatggctgtt agtagtggta cactgatgat gaacaatggc tatgcagtaa 11640
aatcaagact gtagatattg caacagacta taaaattcct ctgtggctta gccaatgtgg 11700
tacttcccac attgtataag aaatttggca agtttagagc aatgtttgaa gtgttgggaa 11760
atttctgtat actcaagagg gcgtttttga caactgtaga acagaggaat caaaaggggg 11820

tgggaggaag ttaaaagaag aggcaggtgc aagagagctt gcagtcccgc tgtgtgtacg 11880

acactggcaa catgaggtct ttgctaatct tggtgctttg cttcctgccc ctggctgcct 11940

taggg 11945

<210> 8
<211> 285
<212> DNA
<213> SV40

<220>
<221> misc_feature
<222> (1)..(285)
<223> SV40 Polyadenylation Sequence

<400> 8
aaagtctaga gtcggggcgg ccggccgctt cgagcagaca tgataagata cattgatgag 60

tttggacaaa ccacaactag aatgcagtga aaaaaatgct ttatttgtga aatttgtgat 120

gctattgctt tatttgtaac cattataagc tgcaataaac aagttaacaa caacaattgc 180

attcatttta tgtttcaggt tcaggggggag gtgtgggagg ttttttaaag caagtaaaac 240

ctctacaaat gtggtaaaat cgataaggat ccgtcgagcg gccgc 285

<210> 9
<211> 5972
<212> DNA
<213> Gallus gallus

<220>
<221> misc_feature
<222> (1)..(5972)
<223> Lysozyme 3prime domain

<400> 9
cgcgtggtag gtggcggggg gttcccagga gagcccccag cgcggacggc agcgccgtca 60

ctcaccgctc cgtctccctc cgcccagggt cgcctggcgc aaccgctgca agggcaccga 120

cgtccaggcg tggatcagag gctgccggct gtgaggagct gccgcgcccg gcccgcccgc 180

tgcacagccg gccgctttgc gagcgcgacg ctacccgctt ggcagtttta aacgcatccc 240

tcattaaaac gactatacgc aaacgccttc ccgtcggtcc gcgtctcttt ccgccgccag 300

ggcgacactc gcggggaggg cgggaagggg gccgggcggg agcccgcggc caaccgtcgc 360

cccgtgacgg caccgccccg cccccgtgac gcggtgcggg cgccggggcc gtggggctga 420

gcgctgcggc ggggccgggc cgggccgggg cgggagctga gcgcggcgcg gctgcgggcg 480

gcgccccctc cggtgcaata tgttcaagag aatggctgag ttcgggcctg actccggggg 540
cagggtgaag gtgcggcgcg ggcggaggga cggggcgggc gcggggccgc ccggcgggtg 600
ccggggcctc tgccggcccg cccggctcgg gctgctgcgg cgcttacggg cgcgcttctc 660
gccgctgccg cttctcttct ctcccgcgca agggcgtcac catcgtgaag ccggtagtgt 720
acgggaacgt ggcgcggtac ttcgggaaga agagggagga ggacgggcac acgcatcagt 780
ggacggttta cgtgaagccc tacaggaacg aggtagggcc cgagcgcgtc ggccgccgtt 840
ctcggagcgc cggagccgtc agcgccgcgc ctgggtgcgc tgtgggacac agcgagcttc 900
tctcgtagga catgtccgcc tacgtgaaaa aaatccagtt caagctgcac gagagctacg 960
ggaatcctct ccgaggtggg tgttgcgtcg gggggtttgc tccgctcggt cccgctgagg 1020
ctcgtcgccc tcatctttct ttcgtgccgc agtcgttacc aaaccgccgt acgagatcac 1080
cgaaacgggc tggggcgaat ttgaaatcat catcaagata tttttcattg atccaaacga 1140
gcgacccgta agtacgctca gcttctcgta gtgcttcccc cgtcctggcg gcccggggct 1200
gggctgctcg ctgctgccgg tcacagtccc gccagccgcg gagctgactg agctcccttt 1260
cccgggacgt gtgctctgtg ttcggtcagc gaggctatcg ggagggcttt ggctgcattt 1320
ggcttctctg gcgcttagcg caggagcacg ttgtgctacg cctgaactac agctgtgaga 1380
aggccgtgga aaccgctctc aaactgattt attggcgaaa tggctctaaa ctaaatcgtc 1440
tcctctcttt ggaaatgctt tagagaaggt ctctgtggta gttcttatgc atctatccta 1500
aagcacttgg ccagacaatt taaagacatc aagcagcatt tatagcaggc acgtttaata 1560
acgaatactg aatttaagta actctgctca cgttgtatga cgtttatttt cgtattcctg 1620
aaagccatta aaatcctgtg cagttgttta gtaagaacag ctgccactgt tttgtatcta 1680
ggagataact ggtgtttccc tacagttctc aagctgataa aactctgtct ttgtatctag 1740
gtaaccctgt atcacttgct gaagcttttt cagtctgaca ccaatgcaat cctgggaaag 1800
aaaactgtag tttctgaatt ctatgatgaa atggtatgaa aattttaatg tcaaccgagc 1860
ctgactttat ttaaaaaaaa ttattgatgg tgctgtgtat tttggtcctt ccttagatat 1920
ttcaagatcc tactgccatg atgcagcaac tgctaacgac gtcccgtcag ctgacacttg 1980
gtgcttacaa gcatgaaaca gagtgtaagt gcaaaatgag gataccttcg ccgaccgtca 2040
ttcactacta atgttttctg tgggatgtga tcgtacagtg agtttggctg tgtgaaattt 2100
gaatagcttg gtattggcag tgatgacgtg atcgatgcct tgcttatcat gtttgaaatg 2160
aagtagaata aatgcagcct gctttatttg agatagtttg gttcatttta tggaatgcaa 2220

gcaaagatta tacttcctca ctgaattgca ctgtccaaag gtgtgaaatg tgtggggatc 2280
tggaggaccg tgaccgaggg acattggatc gctatctccc atttcttttg ctgttaccag 2340
ttcagatttt cttttcacct agtctttaat tcccagggtt ttgtttttc cttggtcata 2400
gtttttgttt ttcactctgg caaatgatgt tgtgaattac actgcttcag ccacaaaact 2460
gatggactga atgaggtcat caaacaaact tttcttcttc cgtatttcct ttttttccc 2520
ccacttatca tttttactgc tgttgttgag tctgtaaggc taaaagtaac tgttttgtgc 2580
tttttcagga cgtgtgcttt ccaaattact gccacatata taaagaaagg ttggaatttt 2640
aaagataatt catgtttctt cttctttttt gccaccacag ttgcagatct tgaagtaaaa 2700
accagggaaa agctggaagc tgccaaaaag aaaaccagtt ttgaaattgc tgagcttaaa 2760
gaaaggttaa aagcaagtcg tgaaaccatc aactgcttaa agagtgaaat cagaaaactc 2820
gaagaggatg atcagtctaa agatatgtga tgagtgttga cttggcaggg agcctataat 2880
gagaatgaaa ggacttcagt cgtggagttg tatgcgttct ctccaattct gtaacggaga 2940
ctgtatgaat ttcatttgca aatcactgca gtgtgtgaca actgactttt tataaatggc 3000
agaaaacaag aatgaatgta tcctcatttt atagttaaaa tctatgggta tgtactggtt 3060
tatttcaagg agaatggatc gtagagactt ggaggccaga ttgctgcttg tattgactgc 3120
atttgagtgg tgtaggaaca ttttgtctat ggtcccgtgt tagtttacag aatgccactg 3180
ttcactgttt tgttttgtat tttacttttt ctactgcaac gtcaaggttt taaaagttga 3240
aaataaaaca tgcaggtttt ttttaaatat ttttttgtct ctatccagtt tgggcttcaa 3300
gtattattgt taacagcaag tcctgattta agtcagaggc tgaagtgtaa tggtattcaa 3360
gatgcttaag tctgttgtca gcaaaacaaa agagaaaact tcataaaatc aggaagttgg 3420
catttctaat aacttcttta tcaacagata agagtttcta gccctgcatc tactttcact 3480
tatgtagttg atgcctttat attttgtgtg tttggatgca ggaagtgatt cctactctgt 3540
tatgtagata ttctatttaa cacttgtact ctgctgtgct tagcctttcc ccatgaaaat 3600
tcagcggctg taaatccccc tcttcttttg tagcctcata cagatggcag accctcaggc 3660
ttataaaggc ttgggcatct tctttactgc tttgagattc tgtgttgcag taacctctgc 3720
cagagaggag aaaagcccca caaacctcat ccccttcttc tatagcaatc agtattacta 3780
atgctttgag aacagagcac tggtttgaaa cgtttgataa ttagcattta acatggcttg 3840
gtaaagatgc agaactgaaa cagctgtgac agtatgaact cagtatggag acttcattaa 3900
gacaaacagc tgttaaaatc aggcatgttt cattgaggag gacggggcaa cttgcaccag 3960

tggtgcccac acaaatcctt cctggcgctg cagaccaatt tttctggcat tctgactgcc 4020
gttgctgctg gtcacagaga gcaactattt ttatcagcca caggcaattt gcttgtagta 4080
ttttccaagt gttgtaggta agtataaatg catcggctcc agagcacttt gagtatactt 4140
attaaaaaca taaatgaaag acaaattagc tttgcttggg tgcacagaac atttttagtt 4200
ccagcctgct ttttggtaga agccctcttc tgaggctaga actgactttg acaagtagag 4260
aaactggcaa cggagctatt gctatcgaag gatccttgtt aacaaagtta atcgtctttt 4320
aaggtttggt ttattcatta aatttgcttt taagctgtag ctgaaaaaga acgtgctgtc 4380
ttccatgcac caggtggcag ctctgtgcaa agtgctctct ggtctcacca gccttttaat 4440
tgccgggatt ctggcacgtc tgagagggct cagactggct tcgtttgttt gaacagcgtg 4500
tactgctttc tgtagacatg gccggtttct ctcctgcagc ttatgaaact gttcacactg 4560
aacacactgg aacaggttgc ccaaggaggc cgtggatgcc ccatccctgg aggcattcaa 4620
ggccaggctg gatgtggctc tgggcagcct ggtctggtgg ttggcgatcc tgcacatagc 4680
agcggggttg aaactcgatg atcactgtgg tccttttcaa cccaggctat tctatgattc 4740
tatgattcaa cagcaaatca tatgtactga gagaggaaac aaacacaagt gctactgttt 4800
gcaagttttg ttcatttggt aaaagagtca ggttttaaaa ttcaaaatct gtctggtttt 4860
ggtgtttttt ttttttatt tattatttct ttggggttct ttttgatgct ttatctttct 4920
ctgccaggac tgtgtgacaa tgggaacgaa aaagaacatg ccaggcactg tcctggattg 4980
cacacgctgg ttgcactcag tagcaggctc agaactgcca gtctttccac agtattactt 5040
tctaaaccta attttaatag cgttagtaga cttccatcac tgggcagtgc ttagtgaatg 5100
ctctgtgtga acgttttact tataagcatg ttggaagttt tgatgttcct ggatgcagta 5160
gggaaggaca gattagctat gtgaaaagta gattctgagt atcggggtta caaaaagtat 5220
agaaacgatg agaaattctt gttgtaacta attggaattt ctttaagcgt tcacttatgc 5280
tacattcata gtatttccat ttaaaagtag gaaaaggtaa aacgtgaaat cgtgtgattt 5340
tcggatggaa caccgccttc ctatgcacct gaccaacttc cagaggaaaa gcctattgaa 5400
agccgagatt aagccaccaa aagaactcat ttgcattgga atatgtagta tttgccctct 5460
tcctcccggg taattactat actttatagg gtgcttatat gttaaatgag tggctggcac 5520
tttttattct cacagctgtg gggaattctg tcctctagga cagaaacaat tttaatctgt 5580
tccactggtg actgctttgt cagcacttcc acctgaagag atcaatacac tcttcaatgt 5640
ctagttctgc aacacttggc aaacctcaca tcttatttca tactctcttc atgcctatgc 5700

ttattaaagc aataatctgg gtaatttttg ttttaatcac tgtcctgacc ccagtgatga 5760

ccgtgtccca cctaaagctc aattcaggtc ctgaatctct tcaactctct atagctaaca 5820

tgaagaatct tcaaaagtta ggtctgaggg acttaaggct aactgtagat gttgttgcct 5880

ggtttctgtg ctgaaggccg tgtagtagtt agagcattca acctctagaa gaagcttggc 5940

cagctggtcg acctgcagat ccggccctcg ag 5972

<210> 10
<211> 18391
<212> DNA
<213> Gallus gallus

<220>
<221> misc_feature
<222> (1)..(237)
<223> 5prime matrix (scaffold) attachment region (MAR)

<220>
<221> misc_feature
<222> (261)..(1564)
<223> 5prime matrix (scaffold) attachment region (MAR)

<220>
<221> misc_feature
<222> (1565)..(1912)
<223> 5prime matrix (scaffold) attachment region (MAR)

<220>
<221> misc_feature
<222> (1930)..(2012)
<223> 5prime matrix (scaffold) attachment region (MAR)

<220>
<221> misc_feature
<222> (2013)..(2671)
<223> Intrinsically curved DNA

<220>
<221> misc_feature
<222> (5848)..(5934)
<223> Transcription enhancer

<220>
<221> misc_feature
<222> (9160)..(9325)
<223> Transcription enhancer

<220>
<221> misc_feature
<222> (9326)..(9626)
<223> Negative regulatory element

<220>
<221> misc_feature
<222> (9621)..(9660)
<223> Hormone response element

<220>
<221> misc_feature
<222> (9680)..(10060)
<223> Hormone response element

<220>
<221> misc_feature
<222> (10576)..(10821)
<223> Chicken CR1 Repeat Sequence

<220>
<221> misc_feature
<222> (10926)..(11193)
<223> Chicken CR1 Repeat Sequence

<220>
<221> misc_feature
<222> (11424)..(11938)
<223> Lysozyme Proximal Promoter and Lysozyme Signal Peptide

<220>
<221> misc_feature
<222> (11946)..(12443)
<223> human interferon alpha 2b codon-optimized for expression in chick
ens

<220>
<221> misc_feature
<222> (12464)..(18391)
<223> Chicken Lysozyme 3prime domain

<400> 10
tgccgccttc tttgatattc actctgttgt atttcatctc ttcttgccga tgaaaggata 60

taacagtctg tataacagtc tgtgaggaaa tacttggtat ttcttctgat cagtgttttt 120

ataagtaatg ttgaatattg gataaggctg tgtgtccttt gtcttgggag acaaagccca 180

cagcaggtgg tggttggggt ggtggcagct cagtgacagg agaggttttt ttgcctgttt 240

```
ttttttttttt ttttttttttt aagtaaggtg ttctttttc ttagtaaatt ttctactgga    300
ctgtatgttt tgacaggtca gaaacatttc ttcaaaagaa gaacctttg gaaactgtac     360
agccctttc tttcattccc tttttgcttt ctgtgccaat gcctttggtt ctgattgcat     420
tatggaaaac gttgatcgga acttgaggtt tttatttata gtgtggcttg aaagcttgga    480
tagctgttgt tacacgagat accttattaa gtttaggcca gcttgatgct ttatttttc     540
cctttgaagt agtgagcgtt ctctggtttt tttcctttga aactggtgag gcttagattt    600
ttctaatggg atttttacc tgatgatcta gttgcatacc caaatgcttg taaatgtttt     660
cctagttaac atgttgataa cttcggattt acatgttgta tatacttgtc atctgtgttt    720
ctagtaaaaa tatatggcat ttatagaaat acgtaattcc tgatttcctt tttttttatc    780
tctatgctct gtgtgtacag gtcaaacaga cttcactcct atttttattt atagaatttt    840
atatgcagtc tgtcgttggt tcttgtgttg taaggataca gccttaaatt tcctagagcg    900
atgctcagta aggcgggttg tcacatgggt tcaaatgtaa aacgggcacg tttggctgct    960
gccttcccga gatccaggac actaaactgc ttctgcactg aggtataaat cgcttcagat   1020
cccagggaag tgcagatcca cgtgcatatt cttaaagaag aatgaatact ttctaaaata   1080
ttttggcata ggaagcaagc tgcatggatt tgtttgggac ttaaattatt ttggtaacgg   1140
agtgcatagg ttttaaacac agttgcagca tgctaacgag tcacagcgtt tatgcagaag   1200
tgatgcctgg atgcctgttg cagctgttta cggcactgcc ttgcagtgag cattgcagat   1260
aggggtgggg tgctttgtgt cgtgttccca cacgctgcca cacagccacc tcccggaaca   1320
catctcacct gctgggtact tttcaaacca tcttagcagt agtagatgag ttactatgaa   1380
acagagaagt tcctcagttg gatattctca tgggatgtct tttttcccat gttgggcaaa   1440
gtatgataaa gcatctctat ttgtaaatta tgcacttgtt agttcctgaa tcctttctat   1500
agcaccactt attgcagcag gtgtaggctc tggtgtggcc tgtgtctgtg cttcaatctt   1560
ttaaagcttc tttggaaata cactgacttg attgaagtct cttgaagata gtaaacagta   1620
cttacctttg atcccaatga aatcgagcat ttcagttgta aaagaattcc gcctattcat   1680
accatgtaat gtaatttac acccccagtg ctgacacttt ggaatatatt caagtaatag    1740
actttggcct caccctcttg tgtactgtat tttgtaatag aaaatatttt aaactgtgca   1800
tatgattatt acattatgaa agagacattc tgctgatctt caaatgtaag aaaatgagga   1860
gtgcgtgtgc ttttataaat acaagtgatt gcaaattagt gcaggtgtcc ttaaaaaaaa   1920
aaaaaaaaag taatataaaa aggaccaggt gttttacaag tgaaatacat tcctatttgg   1980
```

taaacagtta catttttatg aagattacca gcgctgctga ctttctaaac ataaggctgt 2040
attgtcttcc tgtaccattg catttcctca ttcccaattt gcacaaggat gtctgggtaa 2100
actattcaag aaatggcttt gaaatacagc atgggagctt gtctgagttg gaatgcagag 2160
ttgcactgca aaatgtcagg aaatggatgt ctctcagaat gcccaactcc aaaggatttt 2220
atatgtgtat atagtaagca gtttcctgat tccagcaggc caaagagtct gctgaatgtt 2280
gtgttgccgg agacctgtat ttctcaacaa ggtaagatgg tatcctagca actgcggatt 2340
ttaatacatt ttcagcagaa gtacttagtt aatctctacc tttagggatc gtttcatcat 2400
ttttagatgt tatacttgaa atactgcata acttttagct ttcatgggtt cctttttttc 2460
agcctttagg agactgttaa gcaatttgct gtccaacttt tgtgttggtc ttaaactgca 2520
atagtagttt accttgtatt gaagaaataa agaccatttt tatattaaaa aatacttttg 2580
tctgtcttca ttttgacttg tctgatatcc ttgcagtgcc cattatgtca gttctgtcag 2640
atattcagac atcaaaactt aacgtgagct cagtggagtt acagctgcgg ttttgatgct 2700
gttattattt ctgaaactag aaatgatgtt gtcttcatct gctcatcaaa cacttcatgc 2760
agagtgtaag gctagtgaga aatgcataca tttattgata ctttttttaaa gtcaactttt 2820
tatcagattt ttttttcatt tggaaatata ttgttttcta gactgcatag cttctgaatc 2880
tgaaatgcag tctgattggc atgaagaagc acagcactct tcatcttact taaacttcat 2940
tttggaatga aggaagttaa gcaagggcac aggtccatga aatagagaca gtgcgctcag 3000
gagaaagtga acctggattt ctttggctag tgttctaaat ctgtagtgag gaaagtaaca 3060
cccgattcct tgaaagggct ccagctttaa tgcttccaaa ttgaaggtgg caggcaactt 3120
ggccactggt tatttactgc attatgtctc agtttcgcag ctaacctggc ttctccacta 3180
ttgagcatgg actatagcct ggcttcagag gccaggtgaa ggttgggatg ggtggaagga 3240
gtgctgggct gtggctgggg ggactgtggg gactccaagc tgagcttggg gtgggcagca 3300
cagggaaaag tgtgggtaac tatttttaag tactgtgttg caaacgtctc atctgcaaat 3360
acgtagggtg tgtactctcg aagattaaca gtgtgggttc agtaatatat ggatgaattc 3420
acagtggaag cattcaaggg tagatcatct aacgacacca gatcatcaag ctatgattgg 3480
aagcggtatc agaagagcga ggaaggtaag cagtcttcat atgttttccc tccacgtaaa 3540
gcagtctggg aaagtagcac cccttgagca gagacaagga aataattcag gagcatgtgc 3600
taggagaact ttcttgctga attctacttg caagagcttt gatgcctggc ttctggtgcc 3660
ttctgcagca cctgcaaggc ccagagcctg tggtgagctg gagggaaaga ttctgctcaa 3720

gtccaagctt cagcaggtca ttgtctttgc ttcttccccc agcactgtgc agcagagtgg 3780
aactgatgtc gaagcctcct gtccactacc tgttgctgca ggcagactgc tctcagaaaa 3840
agagagctaa ctctatgcca tagtctgaag gtaaaatggg ttttaaaaaa gaaaacacaa 3900
aggcaaaacc ggctgcccca tgagaagaaa gcagtggtaa acatggtaga aaaggtgcag 3960
aagcccccag gcagtgtgac aggcccctcc tgccacctag aggcgggaac aagcttccct 4020
gcctagggct ctgcccgcga agtgcgtgtt tctttggtgg gttttgtttg gcgtttggtt 4080
ttgagattta gacacaaggg aagcctgaaa ggaggtgttg ggcactattt tggtttgtaa 4140
agcctgtact tcaaatatat attttgtgag ggagtgtagc gaattggcca atttaaaata 4200
aagttgcaag agattgaagg ctgagtagtt gagagggtaa cacgtttaat gagatcttct 4260
gaaactactg cttctaaaca cttgtttgag tggtgagacc ttggataggt gagtgctctt 4320
gttacatgtc tgatgcactt gcttgtcctt ttccatccac atccatgcat tccacatcca 4380
cgcatttgtc acttatccca tatctgtcat atctgacata cctgtctctt cgtcacttgg 4440
tcagaagaaa cagatgtgat aatccccagc cgccccaagt ttgagaagat ggcagttgct 4500
tctttccctt tttcctgcta agtaaggatt ttctcctggc tttgacacct cacgaaatag 4560
tcttcctgcc ttacattctg ggcattattt caaatatctt tggagtgcgc tgctctcaag 4620
tttgtgtctt cctactctta gagtgaatgc tcttagagtg aaagagaagg aagagaagat 4680
gttggccgca gttctctgat gaacacacct ctgaataatg gccaaaggtg ggtgggtttc 4740
tctgaggaac gggcagcgtt tgcctctgaa agcaaggagc tctgcggagt tgcagttatt 4800
ttgcaactga tggtggaact ggtgcttaaa gcagattccc taggttccct gctacttctt 4860
ttccttcttg gcagtcagtt tatttctgac agacaaacag ccacccccac tgcaggctta 4920
gaaagtatgt ggctctgcct gggtgtgtta cagctctgcc ctggtgaaag gggattaaaa 4980
cgggcaccat tcatcccaaa caggatcctc attcatggat caagctgtaa ggaacttggg 5040
ctccaacctc aaaacattaa ttggagtacg aatgtaatta aaactgcatt ctcgcattcc 5100
taagtcattt agtctggact ctgcagcatg taggtcggca gctcccactt tctcaaagac 5160
cactgatgga ggagtagtaa aaatggagac cgattcagaa caaccaacgg agtgttgccg 5220
aagaaactga tggaaataat gcatgaattg tgtggtggac attttttttta aatacataaa 5280
ctacttcaaa tgaggtcgga gaaggtcagt gttttattag cagccataaa accaggtgag 5340
cgagtaccat ttttctctac aagaaaaacg attctgagct ctgcgtaagt ataagttctc 5400
catagcggct gaagctcccc cctggctgcc tgccatctca gctggagtgc agtgccattt 5460

ccttggggtt tctctcacag cagtaatggg acaatacttc acaaaaattc tttcttttcc 5520
tgtcatgtgg gatccctact gtgccctcct ggttttacgt taccccctga ctgttccatt 5580
cagcggtttg gaaagagaaa aagaatttgg aaataaaaca tgtctacgtt atcacctcct 5640
ccagcatttt ggtttttaat tatgtcaata actggcttag atttggaaat gagaggggt 5700
tgggtgtatt accgaggaac aaaggaaggc ttatataaac tcaagtcttt tatttagaga 5760
actggcaagc tgtcaaaaac aaaaaggcct taccaccaaa ttaagtgaat agccgctata 5820
gccagcaggg ccagcacgag ggatggtgca ctgctggcac tatgccacgg cctgcttgtg 5880
actctgagag caactgcttt ggaaatgaca gcacttggtg caatttcctt tgtttcagaa 5940
tgcgtagagc gtgtgcttgg cgacagtttt tctagttagg ccacttcttt tttccttctc 6000
tcctcattct cctaagcatg tctccatgct ggtaatccca gtcaagtgaa cgttcaaaca 6060
atgaatccat cactgtagga ttctcgtggt gatcaaatct ttgtgtgagg tctataaaat 6120
atggaagctt atttattttt cgttcttcca tatcagtctt ctctatgaca attcacatcc 6180
accacagcaa attaaaggtg aaggaggctg gtgggatgaa gagggtcttc tagctttacg 6240
ttcttccttg caaggccaca ggaaaatgct gagagctgta gaatacagcc tggggtaaga 6300
agttcagtct cctgctggga cagctaaccg catcttataa ccccttctga gactcatctt 6360
aggaccaaat agggtctatc tggggttttt gttcctgctg ttcctcctgg aaggctatct 6420
cactatttca ctgctcccac ggttacaaac caaagataca gcctgaattt tttctaggcc 6480
acattacata aatttgacct ggtaccaata ttgttctcta tatagttatt tccttcccca 6540
ctgtgtttaa ccccttaagg cattcagaac aactagaatc atagaatggt ttggattgga 6600
aggggcctta aacatcatcc atttccaacc ctctgccatg ggctgcttgc cacccactgg 6660
ctcaggctgc ccagggcccc atccagcctg gccttgagca cctccaggga tggggcaccc 6720
acagcttctc tgggcagcct gtgccaacac ctcaccactc tctgggtaaa gaattctctt 6780
ttaacatcta atctaaatct cttctctttt agtttaaagc cattcctctt tttcccgttg 6840
ctatctgtcc aagaaatgtg tattggtctc cctcctgctt ataagcagga agtactggaa 6900
ggctgcagtg aggtctcccc acagccttct cttctccagg ctgaacaagc ccagctcctt 6960
cagcctgtct tcgtaggaga tcatcttagt ggccctcctc tggacccatt ccaacagttc 7020
cacggctttc ttgtggagcc ccaggtctgg atgcagtact tcagatgggg ccttacaaag 7080
gcagagcaga tggggacaat cgcttacccc tccctgctgg ctgcccctgt tttgatgcag 7140
cccagggtac tgttggcctt tcaggctccc agacccctg ctgatttgtg tcaagctttt 7200

catccaccag aacccacgct tcctggttaa tacttctgcc ctcacttctg taagcttgtt 7260
tcaggagact tccattcttt aggacagact gtgttacacc tacctgccct attcttgcat 7320
atatacattt cagttcatgt ttcctgtaac aggacagaat atgtattcct ctaacaaaaa 7380
tacatgcaga attcctagtg ccatctcagt agggttttca tggcagtatt agcacatagt 7440
caatttgctg caagtacctt ccaagctgcg gcctcccata aatcctgtat ttgggatcag 7500
ttacctttg gggtaagctt ttgtatctgc agagaccctg ggggttctga tgtgcttcag 7560
ctctgctctg ttctgactgc accattttct agatcaccca gttgttcctg tacaacttcc 7620
ttgtcctcca tcctttccca gcttgtatct ttgacaaata caggcctatt tttgtgtttg 7680
cttcagcagc catttaattc ttcagtgtca tcttgttctg ttgatgccac tggaacagga 7740
ttttcagcag tcttgcaaag aacatctagc tgaaaacttt ctgccattca atattcttac 7800
cagttcttct tgtttgaggt gagccataaa ttactagaac ttcgtcactg acaagtttat 7860
gcattttatt acttctatta tgtacttact ttgacataac acagacacgc acatattttg 7920
ctgggatttc cacagtgtct ctgtgtcctt cacatggttt tactgtcata cttccgttat 7980
aaccttggca atctgcccag ctgcccatca caagaaaaga gattcctttt ttattacttc 8040
tcttcagcca ataaacaaaa tgtgagaagc ccaaacaaga acttgtgggg caggctgcca 8100
tcaagggaga gacagctgaa gggttgtgta gctcaataga attaagaaat aataaagctg 8160
tgtcagacag ttttgcctga tttatacagg cacgccccaa gccagagagg ctgtctgcca 8220
aggccacctt gcagtccttg gtttgtaaga taagtcatag gtaacttttc tggtgaattg 8280
cgtggagaat catgatggca gttcttgctg tttactatgg taagatgcta aaataggaga 8340
cagcaaagta acacttgctg ctgtaggtgc tctgctatcc agacagcgat ggcactcgca 8400
caccaagatg agggatgctc ccagctgacg gatgctgggg cagtaacagt gggtcccatg 8460
ctgcctgctc attagcatca cctcagccct caccagccca tcagaaggat catcccaagc 8520
tgaggaaagt tgctcatctt cttcacatca tcaaaccttt ggcctgactg atgcctcccg 8580
gatgcttaaa tgtggtcact gacatcttta tttttctatg atttcaagtc agaacctccg 8640
gatcaggagg gaacacatag tgggaatgta ccctcagctc caaggccaga tcttccttca 8700
atgatcatgc atgctactta ggaaggtgtg tgtgtgtgaa tgtagaattg cctttgttat 8760
tttttcttcc tgctgtcagg aacattttga ataccagaga aaaagaaaag tgctcttctt 8820
ggcatgggag gagttgtcac acttgcaaaa taaaggatgc agtcccaaat gttcataatc 8880
tcagggtctg aaggaggatc agaaactgtg tatacaattt caggcttctc tgaatgcagc 8940

ttttgaaagc tgttcctggc cgaggcagta ctagtcagaa ccctcggaaa caggaacaaa 9000
tgtcttcaag gtgcagcagg aggaaacacc ttgcccatca tgaaagtgaa taaccactgc 9060
cgctgaagga atccagctcc tgtttgagca ggtgctgcac actcccacac tgaaacaaca 9120
gttcattttt ataggacttc caggaaggat cttcttctta agcttcttaa ttatggtaca 9180
tctccagttg gcagatgact atgactactg acaggagaat gaggaactag ctgggaatat 9240
ttctgtttga ccaccatgga gtcacccatt tctttactgg tatttggaaa taataattct 9300
gaattgcaaa gcaggagtta gcgaagatct tcatttcttc catgttggtg acagcacagt 9360
tctggctatg aaagtctgct tacaaggaag aggataaaaa tcatagggat aataaatcta 9420
agtttgaaga caatgaggtt ttagctgcat ttgacatgaa gaaattgaga cctctactgg 9480
atagctatgg tatttacgtg tctttttgct tagttactta ttgaccccag ctgaggtcaa 9540
gtatgaactc aggtctctcg ggctactggc atggattgat tacatacaac tgtaatttta 9600
gcagtgattt agggtttatg agtacttttg cagtaaatca tagggttagt aatgttaatc 9660
tcagggaaaa aaaaaaaaag ccaaccctga cagacatccc agctcaggtg gaaatcaagg 9720
atcacagctc agtgcggtcc cagagaacac agggactctt ctcttaggac ctttatgtac 9780
agggcctcaa gataactgat gttagtcaga agactttcca ttctggccac agttcagctg 9840
aggcaatcct ggaattttct ctccgctgca cagttccagt catcccagtt tgtacagttc 9900
tggcactttt tgggtcaggc cgtgatccaa ggagcagaag ttccagctat ggtcagggag 9960
tgcctgaccg tcccaactca ctgcactcaa acaaaggcga aaccacaaga gtggcttttg 10020
ttgaaattgc agtgtggccc agaggggctg caccagtact ggattgacca cgaggcaaca 10080
ttaatcctca gcaagtgcaa tttgcagcca ttaaattgaa ctaactgata ctacaatgca 10140
atcagtatca acaagtggtt tggcttggaa gatggagtct aggggctcta caggagtagc 10200
tactctctaa tggagttgca ttttgaagca ggacactgtg aaaagctggc ctcctaaaga 10260
ggctgctaaa cattagggtc aattttccag tgcactttct gaagtgtctg cagttcccca 10320
tgcaaagctg cccaaacata gcacttccaa ttgaatacaa ttatatgcag gcgtactgct 10380
tcttgccagc actgtccttc tcaaatgaac tcaacaaaca atttcaaagt ctagtagaaa 10440
gtaacaagct ttgaatgtca ttaaaaagta tatctgcttt cagtagttca gcttatttat 10500
gcccactaga aacatcttgt acaagctgaa cactggggct ccagattagt ggtaaaacct 10560
actttataca atcatagaat catagaatgg cctgggttgg aagggacccc aaggatcatg 10620
aagatccaac acccccgcca caggcagggc caccaacctc cagatctggt actagaccag 10680

gcagcccagg gctccatcca acctggccat gaacacctcc agggatggag catccacaac 10740
ctctctgggc agcctgtgcc agcacctcac caccctctct gtgaagaact tttccctgac 10800
atccaatcta agccttccct ccttgaggtt agatccactc ccccttgtgc tatcactgtc 10860
tactcttgta aaaagttgat tctcctcctt tttggaaggt tgcaatgagg tctccttgca 10920
gccttcttct cttctgcagg atgaacaagc ccagctccct cagcctgtct ttataggaga 10980
ggtgctccag ccctctgatc atctttgtgg ccctcctctg gacccgctcc aagagctcca 11040
catctttcct gtactggggg ccccaggcct gaatgcagta ctccagatgg ggcctcaaaa 11100
gagcagagta aagagggaca atcaccttcc tcaccctgct ggccagccct cttctgatgg 11160
agccctggat acaactggct ttctgagctg caacttctcc ttatcagttc cactattaaa 11220
acaggaacaa tacaacaggt gctgatggcc agtgcagagt ttttcacact tcttcatttc 11280
ggtagatctt agatgaggaa cgttgaagtt gtgcttctgc gtgtgcttct tcctcctcaa 11340
atactcctgc ctgatacctc accccacctg ccactgaatg gctccatggc cccctgcagc 11400
cagggccctg atgaacccgg cactgcttca gatgctgttt aatagcacag tatgaccaag 11460
ttgcacctat gaatacacaa acaatgtgtt gcatccttca gcacttgaga agaagagcca 11520
aatttgcatt gtcaggaaat ggtttagtaa ttctgccaat taaaacttgt ttatctacca 11580
tggctgtttt tatggctgtt agtagtggta cactgatgat gaacaatggc tatgcagtaa 11640
aatcaagact gtagatattg caacagacta taaaattcct ctgtggctta gccaatgtgg 11700
tacttcccac attgtataag aaatttggca agtttagagc aatgtttgaa gtgttgggaa 11760
atttctgtat actcaagagg gcgtttttga caactgtaga acagaggaat caaaaggggg 11820
tgggaggaag ttaaaagaag aggcaggtgc aagagagctt gcagtcccgc tgtgtgtacg 11880
acactggcaa catgaggtct ttgctaatct tggtgctttg cttcctgccc ctggctgcct 11940
tagggtgcga tctgcctcag acccacagcc tgggcagcag gaggaccctg atgctgctgg 12000
ctcagatgag gagaatcagc ctgtttagct gcctgaagga taggcacgat tttggctttc 12060
ctcaagagga gtttggcaac cagtttcaga aggctgagac catccctgtg ctgcacgaga 12120
tgatccagca gatctttaac ctgtttagca ccaaggatag cagcgctgct tgggatgaga 12180
ccctgctgga taagttttac accgagctgt accagcagct gaacgatctg gaggcttgcg 12240
tgatccaggg cgtgggcgtg accgagaccc ctctgatgaa ggaggatagc atcctggctg 12300
tgaggaagta ctttcagagg atcaccctgt acctgaagga gaagaagtac agcccctgcg 12360
cttgggaagt cgtgagggct gagatcatga ggagctttag cctgagcacc aacctgcaag 12420

agagcttgag gtctaaggag taaaaagtct agagtcgggg cggcgcgtgg taggtggcgg 12480
ggggttccca ggagagcccc cagcgcggac ggcagcgccg tcactcaccg ctccgtctcc 12540
ctccgcccag ggtcgcctgg cgcaaccgct gcaagggcac cgacgtccag gcgtggatca 12600
gaggctgccg gctgtgagga gctgccgcgc ccggcccgcc cgctgcacag ccggccgctt 12660
tgcgagcgcg acgctacccg cttggcagtt ttaaacgcat ccctcattaa aacgactata 12720
cgcaaacgcc ttcccgtcgg tccgcgtctc tttccgccgc cagggcgaca ctcgcgggga 12780
gggcgggaag ggggccgggc gggagcccgc ggccaaccgt cgccccgtga cggcaccgcc 12840
ccgcccccgt gacgcggtgc gggcgccggg gccgtggggc tgagcgctgc ggcggggccg 12900
ggccgggccg gggcgggagc tgagcgcggc gcggctgcgg gcggcgcccc ctccggtgca 12960
atatgttcaa gagaatggct gagttcgggc ctgactccgg gggcagggtg aaggtgcggc 13020
gcgggcggag ggacggggcg ggcgcggggc cgcccggcgg gtgccggggc ctctgccggc 13080
ccgcccggct cgggctgctg cggcgcttac gggcgcgctt ctcgccgctg ccgcttctct 13140
tctctcccgc gcaagggcgt caccatcgtg aagccggtag tgtacgggaa cgtggcgcgg 13200
tacttcggga agaagaggga ggaggacggg cacacgcatc agtggacggt ttacgtgaag 13260
ccctacagga acgaggtagg gcccgagcgc gtcggccgcc gttctcggag cgccggagcc 13320
gtcagcgccg cgcctgggtg cgctgtggga cacagcgagc ttctctcgta ggacatgtcc 13380
gcctacgtga aaaaaatcca gttcaagctg cacgagagct acgggaatcc tctccgaggt 13440
gggtgttgcg tcgggggttt tgctccgctc ggtcccgctg aggctcgtcg ccctcatctt 13500
tctttcgtgc cgcagtcgtt accaaaccgc cgtacgagat caccgaaacg ggctggggcg 13560
aatttgaaat catcatcaag atatttttca ttgatccaaa cgagcgaccc gtaagtacgc 13620
tcagcttctc gtagtgcttc ccccgtcctg gcggcccggg gctgggctgc tcgctgctgc 13680
cggtcacagt cccgccagcc gcggagctga ctgagctccc tttcccggga cgtgtgctct 13740
gtgttcggtc agcgaggcta tcgggagggc tttggctgca tttggcttct ctggcgctta 13800
gcgcaggagc acgttgtgct acgcctgaac tacagctgtg agaaggccgt ggaaaccgct 13860
ctcaaactga tttattggcg aaatggctct aaactaaatc gtctcctctc tttggaaatg 13920
ctttagagaa ggtctctgtg gtagttctta tgcatctatc ctaaagcact tggccagaca 13980
atttaaagac atcaagcagc atttatagca ggcacgttta ataacgaata ctgaatttaa 14040
gtaactctgc tcacgttgta tgacgtttat tttcgtattc ctgaaagcca ttaaaatcct 14100
gtgcagttgt ttagtaagaa cagctgccac tgttttgtat ctaggagata actggtgttt 14160

ccctacagtt ctcaagctga taaaactctg tctttgtatc taggtaaccc tgtatcactt 14220
gctgaagctt tttcagtctg acaccaatgc aatcctggga aagaaaactg tagtttctga 14280
attctatgat gaaatggtat gaaaatttta atgtcaaccg agcctgactt tatttaaaaa 14340
aaattattga tggtgctgtg tattttggtc cttccttaga tatttcaaga tcctactgcc 14400
atgatgcagc aactgctaac gacgtcccgt cagctgacac ttggtgctta caagcatgaa 14460
acagagtgta agtgcaaaat gaggatacct tcgccgaccg tcattcacta ctaatgtttt 14520
ctgtgggatg tgatcgtaca gtgagtttgg ctgtgtgaaa tttgaatagc ttggtattgg 14580
cagtgatgac gtgatcgatg ccttgcttat catgtttgaa atgaagtaga ataaatgcag 14640
cctgctttat ttgagatagt ttggttcatt ttatggaatg caagcaaaga ttatacttcc 14700
tcactgaatt gcactgtcca aaggtgtgaa atgtgtgggg atctggagga ccgtgaccga 14760
gggacattgg atcgctatct cccatttctt ttgctgttac cagttcagat tttcttttca 14820
cctagtcttt aattcccagg gttttgtttt ttccttggtc atagtttttg tttttcactc 14880
tggcaaatga tgttgtgaat tacactgctt cagccacaaa actgatggac tgaatgaggt 14940
catcaaacaa acttttcttc ttccgtattt ccttttttttt cccccactta tcatttttac 15000
tgctgttgtt gagtctgtaa ggctaaaagt aactgttttg tgctttttca ggacgtgtgc 15060
tttccaaatt actgccacat atataaagaa aggttggaat tttaaagata attcatgttt 15120
cttcttcttt tttgccacca cagttgcaga tcttgaagta aaaaccaggg aaaagctgga 15180
agctgccaaa aagaaaacca gttttgaaat tgctgagctt aaagaaaggt taaaagcaag 15240
tcgtgaaacc atcaactgct taaagagtga aatcagaaaa ctcgaagagg atgatcagtc 15300
taaagatatg tgatgagtgt tgacttggca gggagcctat aatgagaatg aaaggacttc 15360
agtcgtggag ttgtatgcgt tctctccaat tctgtaacgg agactgtatg aatttcattt 15420
gcaaatcact gcagtgtgtg acaactgact ttttataaat ggcagaaaac aagaatgaat 15480
gtatcctcat tttatagtta aaatctatgg gtatgtactg gtttatttca aggagaatgg 15540
atcgtagaga cttggaggcc agattgctgc ttgtattgac tgcatttgag tggtgtagga 15600
acattttgtc tatggtcccg tgttagttta cagaatgcca ctgttcactg ttttgttttg 15660
tattttactt tttctactgc aacgtcaagg ttttaaaagt tgaaaataaa acatgcaggt 15720
ttttttaaa tatttttttg tctctatcca gtttgggctt caagtattat tgttaacagc 15780
aagtcctgat ttaagtcaga ggctgaagtg taatggtatt caagatgctt aagtctgttg 15840
tcagcaaaac aaaagagaaa acttcataaa atcaggaagt tggcatttct aataacttct 15900

ttatcaacag ataagagttt ctagccctgc atctactttc acttatgtag ttgatgcctt 15960
tatattttgt gtgtttggat gcaggaagtg attcctactc tgttatgtag atattctatt 16020
taacacttgt actctgctgt gcttagcctt tccccatgaa aattcagcgg ctgtaaatcc 16080
ccctcttctt ttgtagcctc atacagatgg cagaccctca ggcttataaa ggcttgggca 16140
tcttctttac tgctttgaga ttctgtgttg cagtaacctc tgccagagag gagaaaagcc 16200
ccacaaacct catccccttc ttctatagca atcagtatta ctaatgcttt gagaacagag 16260
cactggtttg aaacgtttga taattagcat ttaacatggc ttggtaaaga tgcagaactg 16320
aaacagctgt gacagtatga actcagtatg gagacttcat taagacaaac agctgttaaa 16380
atcaggcatg tttcattgag gaggacgggg caacttgcac cagtggtgcc cacacaaatc 16440
cttcctggcg ctgcagacca atttttctgg cattctgact gccgttgctg ctggtcacag 16500
agagcaacta tttttatcag ccacaggcaa tttgcttgta gtattttcca agtgttgtag 16560
gtaagtataa atgcatcggc tccagagcac tttgagtata cttattaaaa acataaatga 16620
aagacaaatt agctttgctt gggtgcacag aacattttta gttccagcct gctttttggt 16680
agaagccctc ttctgaggct agaactgact ttgacaagta gagaaactgg caacggagct 16740
attgctatcg aaggatcctt gttaacaaag ttaatcgtct tttaaggttt ggtttattca 16800
ttaaatttgc ttttaagctg tagctgaaaa agaacgtgct gtcttccatg caccaggtgg 16860
cagctctgtg caaagtgctc tctggtctca ccagcctttt aattgccggg attctggcac 16920
gtctgagagg gctcagactg gcttcgtttg tttgaacagc gtgtactgct ttctgtagac 16980
atggccggtt tctctcctgc agcttatgaa actgttcaca ctgaacacac tggaacaggt 17040
tgcccaagga ggccgtggat gccccatccc tggaggcatt caaggccagg ctggatgtgg 17100
ctctgggcag cctggtctgg tggttggcga tcctgcacat agcagcgggg ttgaaactcg 17160
atgatcactg tggtcctttt caacccaggc tattctatga ttctatgatt caacagcaaa 17220
tcatatgtac tgagagagga aacaaacaca agtgctactg tttgcaagtt ttgttcattt 17280
ggtaaaagag tcaggtttta aaattcaaaa tctgtctggt tttggtgttt ttttttttt 17340
atttattatt tctttggggt tctttttgat gctttatctt tctctgccag gactgtgtga 17400
caatgggaac gaaaaagaac atgccaggca ctgtcctgga ttgcacacgc tggttgcact 17460
cagtagcagg ctcagaactg ccagtctttc cacagtatta ctttctaaac ctaattttaa 17520
tagcgttagt agacttccat cactgggcag tgcttagtga atgctctgtg tgaacgtttt 17580
acttataagc atgttggaag ttttgatgtt cctggatgca gtagggaagg acagattagc 17640

tatgtgaaaa gtagattctg agtatcgggg ttacaaaaag tatagaaacg atgagaaatt 17700

cttgttgtaa ctaattggaa tttctttaag cgttcactta tgctacattc atagtatttc 17760

catttaaaag taggaaaagg taaaacgtga aatcgtgtga ttttcggatg gaacaccgcc 17820

ttcctatgca cctgaccaac ttccagagga aaagcctatt gaaagccgag attaagccac 17880

caaaagaact catttgcatt ggaatatgta gtatttgccc tcttcctccc gggtaattac 17940

tatactttat agggtgctta tatgttaaat gagtggctgg cacttttttat tctcacagct 18000

gtggggaatt ctgtcctcta ggacagaaac aattttaatc tgttccactg gtgactgctt 18060

tgtcagcact tccacctgaa gagatcaata cactcttcaa tgtctagttc tgcaacactt 18120

ggcaaacctc acatcttatt tcatactctc ttcatgccta tgcttattaa agcaataatc 18180

tgggtaattt ttgttttaat cactgtcctg accccagtga tgaccgtgtc ccacctaaag 18240

ctcaattcag gtcctgaatc tcttcaactc tctatagcta acatgaagaa tcttcaaaag 18300

ttaggtctga gggacttaag gctaactgta gatgttgttg cctggtttct gtgctgaagg 18360

ccgtgtagta gttagagcat tcaacctcta g 18391

<210> 11
<211> 586
<212> DNA
<213> Artificial sequence

<220>
<223> MDOT artificial promoter

<400> 11
gtaccgggcc ccccctcgag gtgaatatcc aagaatgcag aactgcatgg aaagcagagc 60

tgcaggcacg atggtgctga gccttagctg cttcctgctg ggagatgtgg atgcagagac 120

gaatgaagga cctgtccctt actcccctca gcattctgtg ctatttaggg ttctaccaga 180

gtccttaaga ggtttttttt ttttttggtc caaaagtctg tttgtttggt tttgaccact 240

gagagcatgt gacacttgtc tcaagctatt aaccaagtgt ccagccaaaa tcgatgtcac 300

aacttgggaa ttttccattt gaagcccctt gcaaaaacaa agagcacctt gcctgctcca 360

gctcctggct gtgaagggtt ttggtgccaa agagtgaaag gcttcctaaa aatgggctga 420

gccggggaag gggggcaact tgggggctat tgagaaacaa ggaaggacaa acagcgttag 480

gtcattgctt ctgcaaacac agccagggct gctcctctat aaaaggggaa gaaagaggct 540

ccgcagccat cacagaccca gaggggacgg tctgtgaatc aagctt 586

<210> 12

<211> 11
<212> PRT
<213> Artificial sequence

<220>
<223> SV40 terminator

<400> 12

Cys Gly Gly Pro Lys Lys Lys Arg Lys Val Gly
1 5 10

<210> 13
<211> 12
<212> DNA
<213> Artificial sequence

<220>
<223> Primer SaltoNotI

<400> 13
tcgagcggcc gc 12

<210> 14
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 14
atggctttga cctttgcctt actggtggct ctcctggtgc tgagctgcaa gagcagctgc 60
tctgtgggct gcgatctgcc tca 83

<210> 15
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 15
gacccacagc ctgggcagca ggaggaccct gatgctgctg gctcagatga ggagaatcag 60
cctgtttagc tgcctgaagg ataggcacga ttttggcttt 100

<210> 16
<211> 62
<212> DNA

<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 16
ctcaagagga gtttggcaac cagtttcaga aggctgagac catccctgtg ctgcacgaga 60
tg 62

<210> 17
<211> 94
<212> DNA
<213> Artificial Sequence

<220>
<223>primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 17
tccagcagat ctttaacctg tttagcacca aggatagcag cgctgcttgg gatgagaccc 60
tgctggataa gttttacacc gagctgtacc agca 94

<210> 18
<211> 77
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 18
ctgaacgatc tggaggcttg cgtgatccag ggcgtgggcg tgaccgagac ccctctgatg 60
aaggaggata gcatcct 77

<210> 19
<211> 82
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 19
gctgtgagga agtactttca gaggatcacc ctgtacctga aggagaagaa gtacagccct 60
tgcgcttggg aagtcgtgag gg 82

<210> 20
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 20
ctgagatcat gaggagcttt agcctgagca ccaacctgca agagagcttg aggtctaagg 60
agtaa 65

<210> 21
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 21
cccaagcttt caccatggct ttgacctttg cctt 34

<210> 22
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 22
atctgcctca gacccacag 19

<210> 23
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 23
gattttggct ttcctcaaga ggagtt 26

<210> 24
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 24
gcacgagatg atccagcaga t 21

<210> 25
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 25
atcgttcagc tgctggtaca 20

<210> 26
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 26
cctcacagcc aggatgctat 20

<210> 27
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 27
atgatctcag ccctcacgac 20

<210> 28
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 28
ctgtgggtct gaggcagat 19

<210> 29
<211> 26
<212> DNA
<213> Artificial Sequence

<220>

<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 29
aactcctctt gaggaaagcc aaaatc 26

<210> 30
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 30
atctgctgga tcatctcgtg c 21

<210> 31
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the formation
of the chicken codon optimized human interferon
2b-encoding nucleic acid

<400> 31
tgctctagac tttttactcc ttagacctca agctct 36

<210> 32
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the synthesis of the MDOT promoter

<400> 32
tcactcgagg tgaatatcca agaat 25

<210> 33
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the synthesis of the MDOT promoter

<400> 33
gagatcgatt ttggctggac acttg 25

<210> 34
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> primer used in the synthesis of the MDOT promoter

<400> 34
cacatcgatg tcacaacttg ggaat 25

<210> 35
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> primer used in the synthesis of the MDOT promoter

<400> 35
tctaagcttc gtcacagacc gtccc 25